



PATENT APPLICATION Docket No.: 2709.1001001

BOX 1 OF 3	Date: 4/4/00	Express Mail Label No. @290720882 US
	. 1 1	Pages Speckenton props 1-1352; Sequence Listing pgs 1-1868 Express Mail Label No. 61290720896 US
BOX 2 OF 3	Date: 4/4/00	
		Pages Sequence Listing, pages 1001-3500
BOX 3 OF 3	Date: 4/4/00	Express Mail Label No. <u>EL290720905 US</u>
	, ,	Pages Sequence Listing, pages 3501-6222
	:	

# TITLE OF THE INVENTION

NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAGILIS FOR DIAGNOSTICS AND THERAPEUTICS

INVENTOR: Gary L. Breton

### RELATED APPLICATIONS:

This application claims the benefit of U.S. Provisional Application Serial Number 60/128,705, filed April 9, 1999, the entire teachings of which are incorporated herein by reference.

# FIELD OF THE INVENTION

The invention relates to isolated nucleic acids and polypeptides derived from *Bacteroides fragilis* that are useful as molecular targets for diagnostics, prophylaxis and treatment of pathological conditions, as well as materials and methods for the diagnosis, prevention, and amelioration of pathological conditions resulting from bacterial infection.

20

## BACKGROUND OF THE INVENTION

The genus *Bacteroides* is a member of the family *Bacteroidaceae*. They are Gram-negative, obligately anaerobic, nonsporeforming rods. The genus contains at least 39 species, and are often isolated from sewage as well as the digestive tract of man, animals, and insects. Bacteroides fragilis was first described in 1898 by Veillon and Zuber, but was called Bacillus fragilis. In 1919, Castellani and Chalmers transferred it to the Bacteroides genus. The "B. fragilis group" refers to the saccharoclastic bacteroids that grow well in bile. Members of this group were previously subspecies of B. fragilis and include B. fragilis, B. distasonis, B. ovatus, B. thetaiotaomicron, and B. vulgatus (Castellani and Chalmers, 1984. Genus I. Bacteroides 1919, 959. Krieg and Holt (editors) In Bergey's Manual of Systematic Bacteriology, 1:604-631).

Bacteroides fragilis accounts for only 1% of the normal flora of the human colon, but is the most common anaerobe isolated from clinical specimens. It is associated with soft tissue infections, abscesses and bacteremia (Moncrief J., et al, 1998. Infect. Immun. 15 66:1735-1739). B. fragilis has also been associated with infection of the skeletal muscle (Katagiri, K., et al, 1996. J. Dermatology. 23:129-132), and meningitis (Aucher, P., et al, 1996. Eur. J. Clin. Microbiol. Infect. Dis. 15:820-823). The B. fragilis group is responsible for 65% of all anaerobic bacteremia cases, with mortality rates in excess of 19% (Redondo, M., et al, 1995. Clinical Infectious Disease. 20:1492-1496).

In 1984, strains of *B. fragilis* were found to cause diarrhea in newborn lambs (Myers, L., et al, 1984. Infect. Immun. 44:241-244). Subsequently, it has been shown that B.fragilis is associated with diarrhea in other livestock and young children. These strains are called enterotoxigenic strains, because they produced a 20KD metalloprotease enterotoxin with intestinal secretory activity (Moncrief J., et al, 1995. Infect. Immun.

63:175-181). 25

> There has been an increase in antibiotic resistance within the Bacteroides fragilis group. While there is still excellent activity of many antibiotics, even some of the most potent agents, the carbapenems and the \( \beta \)-lactamase-inhibitor combinations, are losing

15

20

25

activity (Snydman,D., et al, 1996. Clinical Infectious Diseases. 23:S54-65). The cefoxitin resistance rate has increased from 0% in 1987 to 22% in 1995 (Bianchini, H., et al, 1997. Clinical Infectious Diseases. 25:S268-269). Resistance to metronidazole, co-amoxiclav, and imipenem is rare, but strains have been found that are resistant to one or all of these antibiotics. (Turner,P., et al, 1995.The Lancet. 345:1275-1277). Clindaycin resistance has been shown to be transferred between strains by either plasmid or transposon mechanisms. (Dalmau, D., et al, 1997. Clinical Infectious Diseases. 24:874-877). The increasing resistance to antibiotics commonly used against Bacteroides species may eventually lead to failures of these treatments.

Sequencing and analysis of this genome is crucial for the identification of essential genes for development of drug targets and to reduce the emerging health threat this organism poses.

## SUMMARY OF THE INVENTION

The present invention fulfills the need for diagnostic tools and therapeutics by providing bacterial-specific compositions and methods for detecting *Bacteroides* species including *B. fragilis*, as well as compositions and methods useful for treating and preventing *Bacteroides* infection, in particular, *B. fragilis* infection, in vertebrates including mammals.

The present invention encompasses isolated nucleic acids and polypeptides derived from *B. fragilis* that are useful as reagents for diagnosis of bacterial disease, components of effective antibacterial vaccines, and/or as targets for antibacterial drugs including anti-*B. fragilis* drugs. They can also be used to detect the presence of *B. fragilis* and other *Bacteroides* species in a sample; and in screening compounds for the ability to interfere with the *B. fragilis* life cycle or to inhibit *B. fragilis* infection. They also have use as biocontrol agents for plants.

In one aspect, the invention features compositions of nucleic acids corresponding to entire coding sequences of *B. fragilis* proteins, including surface or secreted proteins or parts thereof, nucleic acids capable of binding mRNA from *B. fragilis* proteins to

15

20

25

block protein translation, and methods for producing *B. fragilis* proteins or parts thereof using peptide synthesis and recombinant DNA techniques. This invention also features antibodies and nucleic acids useful as probes to detect *B. fragilis* infection. In addition, vaccine compositions and methods for the protection or treatment of infection by *B. fragilis* are within the scope of this invention.

The nucleotide sequences provided in SEQ ID NO: 1 - SEQ ID NO: 5222, a fragment thereof, or a nucleotide sequence at least about 99.5% identical to a sequence contained within SEQ ID NO: 1 - SEQ ID NO: 5222 may be "provided" in a variety of medias to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention, i.e., the nucleotide sequence provided in SEQ ID NO: 1 - SEQ ID NO: 5222, a fragment thereof, or a nucleotide sequence at least about 99.5% identical to a sequence contained within SEQ ID NO: 1 - SEQ ID NO: 5222. Uses for and methods for providing nucleotide sequences in a variety of media is well known in the art (see e.g., EPO Publication No. EP 0 756 006).

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any media which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage media, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A person skilled in the art can readily appreciate how any of the presently known computer readable media can be used to create a manufacture comprising computer readable media having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable media. A person skilled in the art can readily adopt any of the presently known methods for recording information on computer readable media to

10

15

generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a person skilled in the art for creating a computer readable media having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable media. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A person skilled in the art can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable media having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide sequence of SEQ ID NO: 1 - SEQ ID NO: 5222, a fragment thereof, or a nucleotide sequence at least about 99.5% identical to SEQ ID NO: 1 - SEQ ID NO: 5222 in computer readable form, a person skilled in the art can routinely access the coding sequence information for a variety of purposes. Computer software is publicly available which allows a person skilled in the art to access sequence information provided in a computer readable media. Examples of such computer software include programs of the "Staden Package", "DNA Star", "MacVector", GCG "Wisconsin Package" (Genetics Computer Group, Madison, WI) and "NCBI Toolbox" (National Center For Biotechnology Information). Suitable programs are described, for example, in Martin J. Bishop, ed., *Guide to Human Genome Computing*, 2d Edition, Academic Press, San Diego, CA. (1998); and Leonard F. Peruski, Jr., and Anne Harwood Peruski, *The Internet and the New Biology: Tools for Genomic and Molecular Research*, American Society for Microbiology, Washington, D.C. (1997).

15

Computer algorithms enable the identification of *B. fragilis* open reading frames (ORFs) within SEQ ID NO: 1 - SEQ ID NO: 5222 which contain homology to ORFs or proteins from other organisms. Examples of such similarity-search algorithms include the BLAST [Altschul et al., J. Mol. Biol. 215:403-410 (1990)] and Smith-Waterman [Smith and Waterman (1981) Advances in Applied Mathematics, 2:482-489] search algorithms. Suitable search algorithms are described, for example, in Martin J. Bishop, ed., *Guide to Human Genome Computing*, 2d Edition, Academic Press, San Diego, CA. (1998); and Leonard F. Peruski, Jr., and Anne Harwood Peruski, *The Internet and the New Biology: Tools for Genomic and Molecular Research*, American Society for Microbiology, Washington, D.C. (1997). Such algorithms are utilized on computer systems as exemplified below. The ORFs so identified represent protein encoding fragments within the *B. fragilis* genome and are useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify commercially important fragments of the *B. fragilis* genome. As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A person skilled in the art can readily appreciate that any one of the currently available computer-based systems is suitable for use in the present invention. The computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means

-15

20

which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the B. fragilis genome which are similar to, or "match", a particular target sequence or target motif. A variety of known algorithms are known in the art and have been disclosed publicly, and a variety of commercially available software for conducting homology-based similarity searches are available and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, FASTA (GCG Wisconsin Package), Bic\_SW (Compugen Bioccelerator), BLASTN2, BLASTP2, BLASTX2 (NCBI) and Motifs (GCG). Suitable software programs are described, for example, in Martin J. Bishop, ed., Guide to Human Genome Computing, 2d Edition, Academic Press. San Diego, CA. (1998); and Leonard F. Peruski, Jr., and Anne Harwood Peruski, The Internet and the New Biology: Tools for Genomic and Molecular Research, American Society for Microbiology, Washington, D.C. (1997). A person skilled in the art can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A person skilled in the art can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that many genes are longer than 500 amino acids, or 1.5 kb in length, and that commercially important fragments of the *B. fragilis* 

15

20

genome, such as sequence fragments involved in gene expression and protein processing, will often be shorter than 30 nucleotides.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a specific functional domain or three-dimensional configuration which is formed upon the folding of the target polypeptide. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic active sites, membrane-spanning regions, and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *B. fragilis* genome possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a person skilled in the art with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *B. fragilis* genome. In the present examples, implementing software which implement the BLASTP2 and bic\_SW algorithms (Altschul et al., J Mol. Biol. 215:403-410 (1990); Compugen Biocellerator) was used to identify open reading frames within the *B. fragilis* genome. A person skilled in the art can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Suitable programs are described, for example, in Martin J. Bishop, ed., *Guide to Human Genome Computing*, 2d Edition, Academic Press, San Diego, CA. (1998); and Leonard F. Peruski, Jr., and Anne Harwood Peruski, *The* 

15

20

25

Internet and the New Biology: Tools for Genomic and Molecular Research, American Society for Microbiology, Washington, D.C. (1997).

The invention features B. fragilis polypeptides, preferably a substantially pure preparation of an B. fragilis polypeptide, or a recombinant B. fragilis polypeptide. In preferred embodiments: the polypeptide has biological activity; the polypeptide has an amino acid sequence at least about 60%, 70%, 80%, 90%, 95%, 98%, or 99% identical to an amino acid sequence of the invention contained in the Sequence Listing, preferably it has about 65% sequence identity with an amino acid sequence of the invention contained in the Sequence Listing, and most preferably it has about 92% to about 99% sequence identity with an amino acid sequence of the invention contained in the Sequence Listing; the polypeptide has an amino acid sequence essentially the same as an amino acid sequence of the invention contained in the Sequence Listing; the polypeptide is at least about 5, 10, 20, 50, 100, or 150 amino acid residues in length; the polypeptide includes at least about 5, preferably at least about 10, more preferably at least about 20, still more preferably at least about 50, 100, or 150 contiguous amino acid residues of the invention contained in the Sequence Listing. In yet another preferred embodiment, the amino acid sequence which differs in sequence identity by about 7% to about 8% from the B. fragilis amino acid sequences of the invention contained in the Sequence Listing is also encompassed by the invention.

In preferred embodiments: the *B. fragilis* polypeptide is encoded by a nucleic acid of the invention contained in the Sequence Listing, or by a nucleic acid having at least about 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a nucleic acid of the invention contained in the Sequence Listing.

In a preferred embodiment, the subject *B. fragilis* polypeptide differs in amino acid sequence at about 1, 2, 3, 5, 10 or more residues from a sequence of the invention contained in the Sequence Listing. The differences, however, are such that the *B. fragilis* polypeptide exhibits an *B. fragilis* biological activity, e.g., the *B. fragilis* polypeptide retains a biological activity of a naturally occurring *B. fragilis* enzyme.

20

25

In preferred embodiments, the polypeptide includes all or a fragment of an amino acid sequence of the invention contained in the Sequence Listing; fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' or 3' to the genomic DNA which encodes a sequence of the invention contained in the Sequence Listing.

In yet other preferred embodiments, the *B. fragilis* polypeptide is a recombinant fusion protein having a first *B. fragilis* polypeptide portion and a second polypeptide portion, e.g., a second polypeptide portion having an amino acid sequence unrelated to *B. fragilis*. The second polypeptide portion can be, e.g., any of glutathione-S-transferase, a DNA binding domain, or a polymerase activating domain. In preferred embodiment the fusion protein can be used in a two-hybrid assay.

Polypeptides of the invention include those which arise as a result of alternative transcription events, alternative RNA splicing events, and alternative translational and postranslational events.

In a preferred embodiment, the encoded *B. fragilis* polypeptide differs (e.g., by amino acid substitution, addition or deletion of at least one amino acid residue) in amino acid sequence at about 1, 2, 3, 5, 10 or more residues, from a sequence of the invention contained in the Sequence Listing. The differences, however, are such that: the *B. fragilis* encoded polypeptide exhibits an *B. fragilis* biological activity, e.g., the encoded *B. fragilis* enzyme retains a biological activity of a naturally occurring *B. fragilis*.

In preferred embodiments, the encoded polypeptide includes all or a fragment of an amino acid sequence of the invention contained in the Sequence Listing; fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' or 3' to the genomic DNA which encodes a sequence of the invention contained in the Sequence Listing.

The *B. fragilis* strain, 14062, from which genomic sequences have been sequenced, has been deposited on July 20, 1998, in the American Type Culture Collection and assigned the ATCC designation # 202158.

15

20

25

Included in the invention are: allelic variations; natural mutants; induced mutants; proteins encoded by DNA that hybridize under high or low stringency conditions to a nucleic acid which encodes a polypeptide of the invention contained in the Sequence Listing (for definitions of high and low stringency see Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6, hereby incorporated by reference); and, polypeptides specifically bound by antisera to *B. fragilis* polypeptides, especially by antisera to an active site or binding domain of *B. fragilis* polypeptide. The invention also includes fragments, preferably biologically active fragments. These and other polypeptides are also referred to herein as *B. fragilis* polypeptide analogs or variants.

The invention further provides nucleic acids, e.g., RNA or DNA and their respective complements, encoding a polypeptide of the invention. This includes double stranded nucleic acids as well as coding and antisense single strands.

In preferred embodiments, the subject *B. fragilis* nucleic acid will include a transcriptional regulatory sequence, e.g., at least one of a transcriptional promoter or transcriptional enhancer sequence, operably linked to the *B. fragilis* gene sequence, e.g., to render the *B. fragilis* gene sequence suitable for expression in a recombinant host cell.

In yet a further preferred embodiment, the nucleic acid which encodes an *B*. *fragilis* polypeptide of the invention, hybridizes under stringent conditions to a nucleic acid probe corresponding to at least about 8 consecutive nucleotides of the invention contained in the Sequence Listing; more preferably to at least about 12 consecutive nucleotides of the invention contained in the Sequence Listing; still more preferably to at least about 20 consecutive nucleotides of the invention contained in the Sequence Listing; most preferably to at least about 40 consecutive nucleotides of the invention contained in the Sequence Listing.

In another aspect, the invention provides a substantially pure nucleic acid having a nucleotide sequence which encodes an *B. fragilis* polypeptide. In preferred embodiments: the encoded polypeptide has biological activity; the encoded polypeptide

15

20

25

has an amino acid sequence at least about 60%, 70%, 80%, 90%, 95%, 98% or 99% homologous to an amino acid sequence of the invention contained in the Sequence Listing; the encoded polypeptide has an amino acid sequence essentially the same as an amino acid sequence of the invention contained in the Sequence Listing; the encoded polypeptide is at least about 5, 10, 20, 50, 100, or 150 amino acids in length; the encoded polypeptide comprises at least about 5, preferably at least about 10, more preferably at least about 20, still more preferably at least about 50, 100, or 150 contiguous amino acids of the invention contained in the Sequence Listing.

In another aspect, the invention encompasses: a vector including a nucleic acid which encodes an *B. fragilis* polypeptide or an *B. fragilis* polypeptide variant as described herein; a host cell transfected with the vector; and a method of producing a recombinant *B. fragilis* polypeptide or *B. fragilis* polypeptide variant; including culturing the cell, e.g., in a cell culture medium, and isolating *an B. fragilis* or *B. fragilis* polypeptide variant, e.g., from the cell or from the cell culture medium.

One embodiment of the invention is directed to substantially isolated nucleic acids. Nucleic acids of the invention include sequences comprising at least about 8 nucleotides in length, more preferably at least about 12 nucleotides in length, even more preferably at least about 15-20 nucleotides in length, that correspond to a subsequence of any one of SEQ ID NO: 1 - SEQ ID NO: 5222 or complements thereof. Alternatively, the nucleic acids comprise sequences contained within any ORF (open reading frame), including a complete protein-coding sequence, of which any of SEQ ID NO: 1 - SEQ ID NO: 5222 forms a part. The invention encompasses sequence-conservative variants and function-conservative variants of these sequences. The nucleic acids may be DNA, RNA, DNA/RNA duplexes, protein-nucleic acid (PNA), or derivatives thereof.

In another aspect, the invention features a purified recombinant nucleic acid having at least about 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% sequence identity or % homology with a sequence of the invention contained in the Sequence Listing

20

The invention also encompasses recombinant DNA (including DNA cloning and expression vectors) comprising these *B. fragilis* -derived sequences; host cells comprising such DNA, including fungal, bacterial, yeast, plant, insect, and mammalian host cells; and methods for producing expression products comprising RNA and polypeptides encoded by the *B. fragilis* sequences. These methods are carried out by incubating a host cell comprising an *B. fragilis* -derived nucleic acid sequence under conditions in which the sequence is expressed. The host cell may be native or recombinant. The polypeptides can be obtained by (a) harvesting the incubated cells to produce a cell fraction and a medium fraction; and (b) recovering the *B. fragilis* polypeptide from the cell fraction, the medium fraction, or both. The polypeptides can also be made by *in vitro* translation.

In another aspect, the invention features nucleic acids capable of binding mRNA of *B. fragilis*. Such nucleic acid is capable of acting as antisense nucleic acid to control the translation of mRNA of *B. fragilis*. A further aspect features a nucleic acid which is capable of binding specifically to an *B. fragilis* nucleic acid. These nucleic acids are also referred to herein as complements and have utility as probes and as capture reagents.

In another aspect, the invention features an expression system comprising an open reading frame corresponding to *B. fragilis* nucleic acid. The nucleic acid further comprises a control sequence compatible with an intended host. The expression system is useful for making polypeptides corresponding to *B. fragilis* nucleic acid.

In another aspect, the invention encompasses: a vector including a nucleic acid which encodes an *B. fragilis* polypeptide or an *B. fragilis* polypeptide variant as described herein; a host cell transfected with the vector; and a method of producing a recombinant *B. fragilis* polypeptide or *B. fragilis* polypeptide variant; including culturing the cell, e.g., in a cell culture medium, and isolating the *B. fragilis* or *B. fragilis* polypeptide variant, e.g., from the cell or from the cell culture medium.

In yet another embodiment of the invention encompasses reagents for detecting bacterial infection, including *B. fragilis* infection, which comprise at least one *B. fragilis* 

15

-derived nucleic acid defined by any one of SEQ ID NO: 1 - SEQ ID NO: 5222, or sequence-conservative or function-conservative variants thereof. Alternatively, the diagnostic reagents comprise nucleotide sequences that are contained within any open reading frames (ORFs), including preferably complete protein-coding sequences, contained within any of SEQ ID NO: 1 - SEQ ID NO: 5222, or polypeptide sequences contained within any of SEQ ID NO: 5223 - SEQ ID NO: 10444, or polypeptides of which any of the above sequences forms a part, or antibodies directed against any of the above peptide sequences or function-conservative variants and/or fragments thereof.

The invention further provides antibodies, preferably monoclonal antibodies, which specifically bind to the polypeptides of the invention. Methods are also provided for producing antibodies in a host animal. The methods of the invention comprise immunizing an animal with at least one *B. fragilis* -derived immunogenic component, wherein the immunogenic component comprises one or more of the polypeptides encoded by any one of SEQ ID NO: 1 - SEQ ID NO: 5222 or sequence-conservative or function-conservative variants thereof; or polypeptides that are contained within any ORFs, including complete protein-coding sequences, of which any of SEQ ID NO: 1 - SEQ ID NO: 5222 forms a part; or polypeptide sequences contained within any of SEQ ID NO: 5223 - SEQ ID NO: 10444; or polypeptides of which any of SEQ ID NO: 5223 - SEQ ID NO: 10444 forms a part. Host animals include any warm blooded animal, including without limitation mammals and birds. Such antibodies have utility as reagents for immunoassays to evaluate the abundance and distribution of *B. fragilis* - specific antigens.

In yet another aspect, the invention provides diagnostic methods for detecting *B*. *fragilis* antigenic components or anti-*B*. *fragilis* antibodies in a sample. *B*. *fragilis* antigenic components may be detected by known processes, including but not limited to detection by a process comprising: (i) contacting a sample suspected to contain a bacterial antigenic component with a bacterial-specific antibody, under conditions in which a stable antigen-antibody complex can form between the antibody and bacterial

15

20

25

antigenic components in the sample; and (ii) detecting any antigen-antibody complex formed in step (i), wherein detection of an antigen-antibody complex indicates the presence of at least one bacterial antigenic component in the sample. In different embodiments of this method, the antibodies used are directed against a sequence encoded by any of SEQ ID NO: 1 - SEQ ID NO: 5222 or sequence-conservative or function-conservative variants thereof, or against a polypeptide sequence contained in any of SEQ ID NO: 5223 - SEQ ID NO: 10444 or function-conservative variants thereof.

In yet another aspect, the invention provides a method for detecting antibacterial-specific antibodies in a sample, which comprises: (i) contacting a sample suspected to contain antibacterial-specific antibodies with an *B. fragilis* antigenic component, under conditions in which a stable antigen-antibody complex can form between the *B. fragilis* antigenic component and antibacterial antibodies in the sample; and (ii) detecting any antigen-antibody complex formed in step (i), wherein detection of an antigen-antibody complex indicates the presence of antibacterial antibodies in the sample. In different embodiments of this method, the antigenic component is encoded by a sequence contained in any of SEQ ID NO: 1 - SEQ ID NO: 5222 or sequence-conservative and function-conservative variants thereof, or is a polypeptide sequence contained in any of SEQ ID NO: 10444 or function-conservative variants thereof.

In another aspect, the invention features a method of generating vaccines for immunizing an individual against *B. fragilis*. The method includes: immunizing a subject with an *B. fragilis* polypeptide, e.g., a surface or secreted polypeptide, or a combination of such peptides or active portion(s) thereof, and a pharmaceutically acceptable carrier. Such vaccines have therapeutic and prophylactic utilities.

In another aspect, the invention features a method of evaluating a compound, e.g., a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *B*. *fragilis* polypeptide. The method includes contacting the compound to be evaluated with an *B*. *fragilis* polypeptide and determining if the compound binds or otherwise interacts with the *B*. *fragilis* polypeptide. Compounds which bind or otherwise interact with *B*.

15

20

25

*fragilis* polypeptides are candidates as modulators, including activators and inhibitors, of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

In another aspect, the invention features a method of evaluating a compound, e.g., a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *B*. *fragilis* nucleic acid, e.g., DNA or RNA. The method includes contacting the compound to be evaluated with an *B*. *fragilis* nucleic acid and determining if the compound binds or otherwise interacts with the *B*. *fragilis* nucleic acid. Compounds which bind *B*. *fragilis* are candidates as modultors, including activators and inhibitors, of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

A particularly preferred embodiment of the invention is directed to a method of screening test compounds for anti-bacterial activity, which method comprises: selecting as a target a bacterial specific sequence, which sequence is essential to the viability of a bacterial species; contacting a test compound with said target sequence; and selecting those test compounds which bind to said target sequence as potential anti-bacterial candidates. In one embodiment, the target sequence selected is specific to a single species, or even a single strain, such as, for example, the strain *B. fragilis* 14062. In a second embodiment, the target sequence is common to at least two species of bacteria. In a third embodiment, the target sequence is common to a family of bacteria. The target sequence may be a nucleic acid sequence or a polypeptide sequence. Methods employing sequences common to more than one species of microorganism may be used to screen candidates for broad spectrum anti-bacterial activity.

The invention also provides methods for preventing or treating disease caused by certain bacteria, including *B. fragilis*, which are carried out by administering to an animal in need of such treatment, in particular a warm-blooded vertebrate, including but not limited to birds and mammals, a compound that specifically inhibits or interferes with the function of a bacterial polypeptide or nucleic acid. In a particularly preferred embodiment, the mammal to be treated is human.

## DETAILED DESCRIPTION OF THE INVENTION

The sequences of the present invention include the specific nucleic acid and amino acid sequences set forth in the Sequence Listing that forms a part of the present specification, and which are designated SEQ ID NO: 1 - SEQ ID NO: 10444. Use of the terms "SEQ ID NO: 1 - SEQ ID NO: 5222 ", " SEQ ID NO: 5223 - SEQ ID NO: 10444, "the sequences depicted in Table 2", etc., is intended, for convenience, to refer to each individual SEQ ID NO *individually*, and is not intended to refer to the genus of these sequences unless such reference would be indicated. In other words, it is a shorthand for listing all of these sequences individually. The invention encompasses each sequence individually, as well as any combination thereof.

### DEFINITIONS

10

15

20

"Nucleic acid" or "polynucleotide" as used herein refers to purine- and pyrimidine-containing polymers of any length, either polyribonucleotides or polydeoxyribonucleotides or mixed polyribo-polydeoxyribo nucleotides. This includes single- and double-stranded molecules, i.e., DNA-DNA, DNA-RNA and RNA-RNA hybrids, as well as "protein nucleic acids" (PNA) formed by conjugating bases to an amino acid backbone. This also includes nucleic acids containing modified bases.

A nucleic acid or polypeptide sequence that is "derived from" a designated sequence refers to a sequence that corresponds to a region of the designated sequence. For nucleic acid sequences, this encompasses sequences that are homologous or complementary to the sequence, as well as "sequence-conservative variants" and "function-conservative variants." For polypeptide sequences, this encompasses "function-conservative variants." Sequence-conservative variants are those in which a change of one or more nucleotides in a given codon position results in no alteration in the amino acid encoded at that position. Function-conservative variants are those in which a given amino acid residue in a polypeptide has been changed without altering the overall conformation and function of the native polypeptide, including, but not limited to,

20

25

replacement of an amino acid with one having similar physico-chemical properties (such as, for example, acidic, basic, hydrophobic, and the like). "Function-conservative" variants also include any polypeptides that have the ability to elicit antibodies specific to a designated polypeptide.

An "B. fragilis -derived" nucleic acid or polypeptide sequence may or may not be present in other bacterial species, and may or may not be present in all B. fragilis strains. This term is intended to refer to the source from which the sequence was originally isolated. Thus, an B. fragilis -derived polypeptide, as used herein, may be used, e.g., as a target to screen for a broad spectrum antibacterial agent, to search for homologous proteins in other species of bacteria or in eukaryotic organisms such asbacteria humans, etc.

A purified or isolated polypeptide or a substantially pure preparation of a polypeptide are used interchangeably herein and, as used herein, mean a polypeptide that has been separated from other proteins, lipids, and nucleic acids with which it naturally occurs. Preferably, the polypeptide is also separated from substances, e.g., antibodies or gel matrix, e.g., polyacrylamide, which are used to purify it. Preferably, the polypeptide constitutes at least about 10, 20, 50 70, 80 or 95% dry weight of the purified preparation. Preferably, the preparation contains sufficient polypeptide to allow protein sequencing; at least about 1, 10, or preferably 100 mg of polypeptide.

A purified preparation of cells refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least about 10%, more preferably at least about 50%, of the subject cells.

A purified or isolated or a substantially pure nucleic acid, e.g., a substantially pure DNA, (are terms used interchangeably herein) is a nucleic acid which is one or both of the following: not immediately contiguous with both of the coding sequences with which it is immediately contiguous (i.e., one at the 5' end and one at the 3' end) in the naturally-occurring genome of the organism from which the nucleic acid is derived; or

15

20

which is substantially free of a nucleic acid with which it occurs in the organism from which the nucleic acid is derived. The term includes, for example, a recombinant DNA which is incorporated into a vector, e.g., into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other DNA sequences. Substantially pure DNA also includes a recombinant DNA which is part of a hybrid gene encoding additional *B. fragilis* DNA sequence.

A "contig" as used herein is a nucleic acid representing a continuous stretch of genomic sequence of an organism.

An "open reading frame", also referred to herein as ORF, is a region of nucleic acid which encodes a polypeptide. This region may represent a portion of a coding sequence or a total sequence and can be determined from a stop to stop codon or from a start to stop codon.

As used herein, a "coding sequence" is a nucleic acid which is transcribed into messenger RNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the five prime terminus and a translation stop code at the three prime terminus. A coding sequence can include but is not limited to messenger RNA, synthetic DNA, and recombinant nucleic acid sequences.

A "complement" of a nucleic acid as used herein refers to an anti-parallel or antisense sequence that participates in Watson-Crick base-pairing with the original sequence.

A "gene product" is a protein or structural RNA which is specifically encoded by a gene.

As used herein, the term "probe" refers to a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest. Probes are often associated with or capable of associating with a label. A label is a chemical moiety

25

capable of detection. Typical labels comprise dyes, radioisotopes, luminescent and chemiluminescent moieties, fluorophores, enzymes, precipitating agents, amplification sequences, and the like. Similarly, a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest and immobilizes such molecule is referred herein as a "capture ligand". Capture ligands are typically associated with or capable of associating with a support such as nitro-cellulose, glass, nylon membranes, beads, particles and the like. The specificity of hybridization is dependent on conditions such as the base pair composition of the nucleotides, and the temperature and salt concentration of the reaction. These conditions are readily discernable to one of ordinary skill in the art using routine experimentation.

"Homologous" refers to the sequence similarity or sequence identity between two polypeptides or between two nucleic acid molecules. When a position in both of the two compared sequences is occupied by the same base or amino acid monomer subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then the molecules are homologous at that position. The percent of homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences divided by the number of positions compared x 100. For example, if 6 of 10 of the positions in two sequences are matched or homologous then the two sequences are 60% homologous. By way of example, the DNA sequences ATTGCC and TATGGC share 50% homology. Generally, a comparison is made when two sequences are aligned to give maximum homology.

Nucleic acids are hybridizable to each other when at least one strand of a nucleic acid can anneal to the other nucleic acid under defined stringency conditions. Stringency of hybridization is determined by: (a) the temperature at which hybridization and/or washing is performed; and (b) the ionic strength and polarity of the hybridization and washing solutions. Hybridization requires that the two nucleic acids contain complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. Typically, hybridization of two sequences at high

15

20

25

stringency (such as, for example, in a solution of 0.5X SSC, at 65° C) requires that the sequences be essentially completely homologous. Conditions of intermediate stringency (such as, for example, 2X SSC at 65° C) and low stringency (such as, for example 2X SSC at 55° C) require correspondingly less overall complementarity between the hybridizing sequences. (1X SSC is 0.15 M NaCl, 0.015 M Na citrate).

The terms peptides, proteins, and polypeptides are used interchangeably herein.

As used herein, the term "surface protein" refers to all surface accessible proteins, e.g. inner and outer membrane proteins, proteins adhering to the cell wall, and secreted proteins.

A polypeptide has *B. fragilis* biological activity if it has one, two or preferably more of the following properties: (1) if when expressed in the course of an *B. fragilis* infection, it can promote, or mediate the attachment of *B. fragilis* to a cell; (2) it has an enzymatic activity, structural or regulatory function characteristic of an *B. fragilis* protein; (3) the gene which encodes it can rescue a lethal mutation in an *B. fragilis* gene. A polypeptide has biological activity if it is an antagonist, agonist, or super-agonist of a polypeptide having one of the above-listed properties.

A biologically active fragment or analog is one having an *in vivo* or *in vitro* activity which is characteristic of the *B. fragilis* polypeptides of the invention contained in the Sequence Listing, or of other naturally occurring *B. fragilis* polypeptides, e.g., one or more of the biological activities described herein. Especially preferred are fragments which exist *in vivo*, e.g., fragments which arise from post transcriptional processing or which arise from translation of alternatively spliced RNA's. Fragments include those expressed in native or endogenous cells as well as those made in expression systems, e.g., in CHO (Chinese Hamster Ovary) cells. Because peptides such as *B. fragilis* polypeptides often exhibit a range of physiological properties and because such properties may be attributable to different portions of the molecule, a useful *B. fragilis* fragment or *B. fragilis* analog is one which exhibits a biological activity in any biological assay for *B. fragilis* activity. The fragment or analog possesses about 10%, preferably

15

about 40%, more preferably about 60%, 70%, 80% or 90% or greater of the activity of *B. fragilis*, in any *in vivo* or *in vitro* assay.

Analogs can differ from naturally occurring *B. fragilis* polypeptides in amino acid sequence or in ways that do not involve sequence, or both. Non-sequence modifications include changes in acetylation, methylation, phosphorylation, carboxylation, or glycosylation. Preferred analogs include *B. fragilis* polypeptides (or biologically active fragments thereof) whose sequences differ from the wild-type sequence by one or more conservative amino acid substitutions or by one or more non-conservative amino acid substitutions, deletions, or insertions which do not substantially diminish the biological activity of the *B. fragilis* polypeptide. Conservative substitutions typically include the substitution of one amino acid for another with similar characteristics, e.g., substitutions within the following groups: valine, glycine; glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Other conservative substitutions can be made in view of the table below.

TABLE 1
CONSERVATIVE AMINO ACID REPLACEMENTS

For Amino Acid	Code	Replace with any of
Alanine	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile,
		D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	С	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	Е	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, β-Ala, Acp
Isoleucine	I	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met

Leucine	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, cis-3,4, or 5-phenylproline
Proline	P	D-Pro, L-I-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

Other analogs within the invention are those with modifications which increase peptide stability; such analogs may contain, for example, one or more non-peptide bonds (which replace the peptide bonds) in the peptide sequence. Also included are: analogs that include residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g.,  $\beta$  or  $\gamma$  amino acids; and cyclic analogs.

As used herein, the term "fragment", as applied to an *B. fragilis* analog, will ordinarily be at least about 20 residues, more typically at least about 40 residues, preferably at least about 60 residues in length. Fragments of *B. fragilis* polypeptides can be generated by methods known to those skilled in the art. The ability of an *Bacteroides* fragment to exhibit a biological activity of *B. fragilis* polypeptide can be assessed by methods known to those skilled in the art as described herein. Also included are *B. fragilis* polypeptides containing residues that are not required for biological activity of the peptide or that result from alternative mRNA splicing or alternative protein processing events.

15

25

An "immunogenic component" as used herein is a moiety, such as an *B. fragilis* polypeptide, analog or fragment thereof, that is capable of eliciting a humoral and/or cellular immune response in a host animal.

An "antigenic component" as used herein is a moiety, such as an *B. fragilis* polypeptide, analog or fragment thereof, that is capable of binding to a specific antibody with sufficiently high affinity to form a detectable antigen-antibody complex.

The term "antibody" as used herein is intended to include fragments thereof which are specifically reactive with *B. fragilis* polypeptides.

As used herein, the term "cell-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well.

Misexpression, as used herein, refers to a non-wild type pattern of gene expression. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of increased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

As used herein, "host cells" and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refers to cells which can

15

20

25

become or have been used as recipients for a recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood by individuals skilled in the art that the progeny of a single parental cell may not necessarily be completely identical in genomic or total DNA compliment to the original parent, due to accident or deliberate mutation.

As used herein, the term "control sequence" refers to a nucleic acid having a base sequence which is recognized by the host organism to effect the expression of encoded sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include a promoter, ribosomal binding site, terminators, and in some cases operators; in eukaryotes, generally such control sequences include promoters, terminators and in some instances, enhancers. The term control sequence is intended to include at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

As used herein, the term "operably linked" refers to sequences joined or ligated to function in their intended manner. For example, a control sequence is operably linked to coding sequence by ligation in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence and host cell.

The "metabolism" of a substance, as used herein, means any aspect of the expression, function, action, or regulation of the substance. The metabolism of a substance includes modifications, e.g., covalent or non-covalent modifications of the substance. The metabolism of a substance includes modifications, e.g., covalent or non-covalent modification, the substance induces in other substances. The metabolism of a substance also includes changes in the distribution of the substance. The metabolism of a substance includes changes the substance induces in the distribution of other substances.

A "sample" as used herein refers to a biological sample, such as, for example, tissue or fluid isloated from an individual (including without limitation plasma, serum,

cerebrospinal fluid, lymph, tears, saliva and tissue sections) or from *in vitro* cell culture constituents, as well as samples from the environment.

Technical and scientific terms used herein have the meanings commonly understood by one of ordinary skill in the art to which the present invention pertains, unless otherwise defined. Reference is made herein to various methodologies known to those of skill in the art. Publications and other materials setting forth such known methodologies to which reference is made are incorporated herein by reference in their entireties as though set forth in full. The practice of the invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Sambrook, Fritsch, and Maniatis, Molecular Cloning; Laboratory Manual 2nd ed. (1989); DNA Cloning, Volumes I and II (D.N Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed, 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); the series, Methods in Enzymology (Academic Press, Inc.), particularly Vol. 154 and Vol. 155 (Wu and Grossman, eds.); PCR-A Practical Approach (McPherson, Quirke, and Taylor, eds., 1991); Immunology, 2d Edition, 1989, Roitt et al., C.V. Mosby Company, and New York; Advanced Immunology, 2d Edition, 1991, Male et al., Grower Medical Publishing, New York.; DNA Cloning: A Practical Approach, Volumes I and II, 1985 (D.N. Glover ed.); Oligonucleotide Synthesis, 1984, (M.L. Gait ed); Transcription and Translation, 1984 (Hames and Higgins eds.); *Animal Cell Culture*, 1986 (R.I. Freshney ed.); Immobilized Cells and Enzymes, 1986 (IRL Press); Perbal, 1984, A Practical Guide to Molecular Cloning; Gene Transfer Vectors for Mammalian Cells, 1987 (J. H. Miller and M. P. Calos eds., Cold Spring Harbor Laboratory); Martin J. Bishop, ed., Guide to Human Genome Computing, 2d Edition, Academic Press, San Diego, CA. (1998); and Leonard F. Peruski, Jr., and Anne Harwood Peruski, The Internet and the New Biology: Tools for Genomic and Molecular Research, American Society for Microbiology, Washington, D.C. (1997).

15

25

Any suitable materials and/or methods known to those of skill can be utilized in carrying out the present invention; however, preferred materials and/or methods are described. Materials, reagents and the like to which reference is made in the following description and examples are obtainable from commercial sources, unless otherwise noted.

# B. FRAGILIS GENOMIC SEQUENCE

This invention provides nucleotide sequences of the genome of *B. fragilis* which thus comprises a DNA sequence library of *B. fragilis* genomic DNA. The detailed description that follows provides nucleotide sequences of *B. fragilis*, and also describes how the sequences were obtained and how ORFs and protein-coding sequences were identified. Also described are compositions and methods of using the disclosed *B. fragilis* sequences in methods including diagnostic and therapeutic applications. Furthermore, the library can be used as a database for identification and comparison of medically important sequences in this and other strains of *B. fragilis*.

To determine the genomic sequence of *B. fragilis*, DNA from strain 14062 of *B. fragilis* was isolated after Zymolyase digestion, sodium dodecyl sulfate lysis, potassium acetate precipitation, phenol:chloroform extractionand ethanol precipitation (Soll, D.R., T. Srikantha and S.R. Lockhart: Characterizing Developmentally Regulated Genes in *B. fragilis*. In Microbial Genome Methods. K.W. Adolph, editor. CRC Press. New York. p 17-37.). DNA was sheared hydrodynamically using an HPLC (Oefner, et. al., 1996) to an insert size of 2000-3000 bp. After size fractionation by gel electrophoresis the fragments were blunt-ended, ligated to adapter oligonucleotides and cloned into the pGTC (Thomann) vector to construct a "shotgun" subclone library.

DNA sequencing was achieved using established ABI sequencing methods on ABI377 automated DNA sequencers. The cloning and sequencing procedures are described in more detail in the Exemplification.

Individual sequence reads were assembled using PHRAP (P. Green, Abstracts of DOE Human Genome Program Contractor-Grantee Workshop V, Jan. 1996, p.157). The average contig length was about 3-4 kb.

All subsequent steps were based on sequencing by ABI377 automated DNA sequencing methods. The cloning and sequencing procedures are described in more detail in the Exemplification.

A variety of approaches may be used to order the contigs so as to obtain a continuous sequence representing the entire *B. fragilis* genome. Synthetic oligonucleotides are designed that are complementary to sequences at the end of each contig. These oligonucleotides may be hybridized to libaries of *B. fragilis* genomic DNA in, for example, lambda phage vectors or plasmid vectors to identify clones that contain sequences corresponding to the junctional regions between individual contigs. Such clones are then used to isolate template DNA and the same oligonucleotides are used as primers in polymerase chain reaction (PCR) to amplify junctional fragments, the nucleotide sequence of which is then determined.

The *B. fragilis* sequences were analyzed for the presence of open reading frames (ORFs) comprising at least 180 nucleotides. As a result of the analysis of ORFs based on stop-to-stop codon reads, it should be understood that these ORFs may not correspond to the ORF of a naturally-occurring *B. fragilis* polypeptide. These ORFs may contain start codons which indicate the initiation of protein synthesis of a naturally-occurring *B. fragilis* polypeptide. Such start codons within the ORFs provided herein were identified by those of ordinary skill in the relevant art, and the resulting ORF and the encoded *B. fragilis* polypeptide is within the scope of this invention. For example, within the ORFs a codon such as AUG or GUG (encoding methionine or valine) which is part of the initiation signal for protein synthesis were identified and the portion of an ORF to corresponding to a naturally-occurring *B. fragilis* polypeptide was recognized. The predicted coding regions were defined by evaluating the coding potential of such

15

20

25

sequences with the program GENEMARK<sup>TM</sup> (Borodovsky and McIninch, 1993, *Comp.* . 17:123).

Each predicted ORF amino acid sequence was compared with all sequences found in current GENBANK, SWISS-PROT, and PIR databases using the BLAST algorithm. BLAST identifies local alignments occurring by chance between the ORF sequence and the sequence in the databank (Altschal et al., 1990, L Mol. Biol. 215:403-410). Homologous ORFs (probabilities less than 10<sup>-5</sup> by chance) and ORF's that are probably non-homologous (probabilities greater than 10<sup>-5</sup> by chance) but have good codon usage were identified. Both homologous, sequences and non-homologous sequences with good codon usage, are likely to encode proteins and are encompassed by the invention.

## B. FRAGILIS NUCLEIC ACIDS

The present invention provides a library of *B. fragilis* -derived nucleic acid sequences. The libraries provide probes, primers, and markers which are used as markers in epidemiological studies. The present invention also provides a library of *B. fragilis* - derived nucleic acid sequences which comprise or encode targets for therapeutic drugs.

The nucleic acids of this invention may be obtained directly from the DNA of the above referenced *B. fragilis* strain by using the polymerase chain reaction (PCR). See "PCR, A Practical Approach" (McPherson, Quirke, and Taylor, eds., IRL Press, Oxford, UK, 1991) for details about the PCR. High fidelity PCRis used to ensure a faithful DNA copy prior to expression. In addition, the authenticity of amplified products is verified by conventional sequencing methods. Clones carrying the desired sequences described in this invention may also be obtained by screening the libraries by means of the PCR or by hybridization of synthetic oligonucleotide probes to filter lifts of the library colonies or plaques as known in the art (see, e.g., Sambrook et al., Molecular Cloning, A Laboratory Manual 2nd edition, 1989, Cold Spring Harbor Press, NY).

It is also possible to obtain nucleic acids encoding *B. fragilis* polypeptides from a cDNA library in accordance with protocols herein described. A cDNA encoding an *B. fragilis* polypeptide can be obtained by isolating total mRNA from an appropriate strain. Double stranded cDNAs can then be prepared from the total mRNA. Subsequently, the cDNAs can be inserted into a suitable plasmid or viral (e.g., bacteriophage) vector using any one of a number of known techniques. Genes encoding *B. fragilis* polypeptides can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acids of the invention can be DNA or RNA. Preferred nucleic acids of the invention are contained in the Sequence Listing.

The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent

In another example, DNA can be chemically synthesized using, e.g., the phosphoramidite solid support method of Matteucci *et al.*, 1981, *J. Am. Chem. Soc.* 103:3185, the method of Yoo *et al.*, 1989, *J. Biol. Chem.* 764:17078, or other well known methods. This can be done by sequentially linking a series of oligonucleotide cassettes comprising pairs of synthetic oligonucleotides, as described below.

Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

Nucleic acids isolated or synthesized in accordance with features of the present invention are useful, by way of example, without limitation, as probes, primers, capture ligands, antisense genes and for developing expression systems for the synthesis of proteins and peptides corresponding to such sequences. As probes, primers, capture ligands and antisense agents, the nucleic acid normally consists of all or part (approximately twenty or more nucleotides for specificity as well as the ability to form

10

20

stable hybridization products) of the nucleic acids of the invention contained in the Sequence Listing. These uses are described in further detail below.

### **PROBES**

A nucleic acid isolated or synthesized in accordance with the sequence of the invention contained in the Sequence Listing can be used as a probe to specifically detect *B. fragilis*. With the sequence information set forth in the present application, sequences of twenty or more nucleotides are identified which provide the desired inclusivity and exclusivity with respect to *B. fragilis*, and extraneous nucleic acids likely to be encountered during hybridization conditions. More preferably, the sequence will comprise at least about twenty to thirty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules.

Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques. Individuals skilled in the art will readily recognize that the nucleic acids, for use as probes, can be provided with a label to facilitate detection of a hybridization product.

Nucleic acid isolated and synthesized in accordance with the sequence of the invention contained in the Sequence Listing can also be useful as probes to detect homologous regions (especially homologous genes) of other *Bacteroides* species using appropriate stringency hybridization conditions as described herein.

### **CAPTURE LIGAND**

For use as a capture ligand, the nucleic acid selected in the manner described above with respect to probes, can be readily associated with a support. The manner in which nucleic acid is associated with supports is well known. Nucleic acid having twenty or more nucleotides in a sequence of the invention contained in the Sequence Listing have utility to separate *B. fragilis* nucleic acid from one strain from the nucleic acid of other another strain as well as from other organisms. Nucleic acid having twenty

15

20

25

or more nucleotides in a sequence of the invention contained in the Sequence Listing can also have utility to separate other *Bacteroides* species from each other and from other organisms. Preferably, the sequence will comprise at least about twenty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules. Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques.

# **PRIMERS**

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as primers for the amplification of B. fragilis nucleic acid. These nucleic acids may also have utility as primers for the amplification of nucleic acids in other Bacteroides species. With respect to polymerase chain reaction (PCR) techniques, nucleic acid sequences of  $\geq 10$ -15 nucleotides of the invention contained in the Sequence Listing have utility in conjunction with suitable enzymes and reagents to create copies of B. fragilis nucleic acid. More preferably, the sequence will comprise twenty or more nucleotides to convey stability to the hybridization product formed between the primer and the intended target molecules. Binding conditions of primers greater than 100 nucleotides are more difficult to control to obtain specificity. High fidelity PCR can be used to ensure a faithful DNA copy prior to expression. In addition, amplified products can be checked by conventional sequencing methods.

The copies can be used in diagnostic assays to detect specific sequences, including genes from *B. fragilis* and/or other *Bacteroides* species. The copies can also be incorporated into cloning and expression vectors to generate polypeptides corresponding to the nucleic acid synthesized by PCR, as is described in greater detail herein.

The nucleic acids of the present invention find use as templates for the recombinant production of *B. fragilis* -derived peptides or polypeptides

20

25



### **ANTISENSE**

Nucleic acid or nucleic acid-hybridizing derivatives isolated or synthesized in accordance with the sequences described herein have utility as antisense agents to prevent the expression of *B. fragilis* genes. These sequences also have utility as antisense agents to prevent expression of genes of other *Bacteroides* species.

In one embodiment, nucleic acid or derivatives corresponding to *B. fragilis* nucleic acids is loaded into a suitable carrier such as a liposome or bacteriophage for introduction into bacterial cells. For example, a nucleic acid having twenty or more nucleotides is capable of binding to bacteria nucleic acid or bacteria messenger RNA. Preferably, the antisense nucleic acid is comprised of 20 or more nucleotides to provide necessary stability of a hybridization product of non-naturally occurring nucleic acid and bacterial nucleic acid and/or bacterial messenger RNA. Nucleic acid having a sequence greater than 1000 nucleotides in length is difficult to synthesize but can be generated by recombinant DNA techniques. Methods for loading antisense nucleic acid in liposomes is known in the art as exemplified by U.S. Patent 4,241,046 issued December 23, 1980 to Papahadjopoulos et al.

The present invention encompasses isolated polypeptides and nucleic acids derived from *B. fragilis* that are useful as reagents for diagnosis of bacterial infection, components of effective anti-bacterial vaccines, and/or as targets for anti-bacterial drugs, including anti-*B. fragilis* drugs.

## EXPRESSION OF B. FRAGILIS NUCLEIC ACIDS

Table 2, which is appended herewith and which forms part of the present specification, provides a list of open reading frames (ORFs) in both strands and a putative identification of the particular function of a polypeptide which is encoded by each ORF, based on the homology match (determined by the BLASTP2 algorithm) of the predicted polypeptide with known proteins encoded by ORFs in other organisms. An ORF is a region of nucleic acid which encodes a polypeptide. This region may represent

a portion of a coding sequence or a total sequence and was determined from stop to stop codons. The first column contains a designation for the ORF ("ORF Name"). The second and third columns list the SEQ ID numbers for the nucleic acid ("NT ID") and amino acid ("AA ID") sequences corresponding to each ORF, respectively. The fourth and fifth columns list the length of the nucleic acid ORF ("NT Length") and the length of the amino acid ORF ("AA Length"), respectively. The nucleotide sequence corresponding to each ORF begins at the first nucleotide immediately following a stop codon and ends at the nucleotide immediately preceding the next downstream stop codon in the same reading frame. It will be recognized by one skilled in the art that the natural translation initiation sites will correspond to ATG, GTG, or TTG codons located within the ORFs. The natural initiation sites depend not only on the sequence of a start codon but also on the context of the DNA sequence adjacent to the start codon. Usually, a recognizable ribosome binding site is found within 20 nucleotides upstream from the initiation codon. In some cases where genes are translationally coupled and coordinately expressed together in "operons", ribosome binding sites are not present, but the initiation codon of a downstream gene may occur very close to, or overlap, the stop codon of the an upstream gene in the same operon. The correct start codons can be generally identified without undue experimentation because only a few codons need be tested. It is recognized that the translational machinery in bacteria initiates all polypeptide chains with the amino acid methionine, regardless of the sequence of the start codon. In some cases, polypeptides are post-translationally modified, resulting in an N-terminal amino acid other than methionine in vivo. The sixth and seventh columns provide metrics for assessing the likelihood of the homology match (determined by the BLASTP2 algorithm), as is known in the art, to the genes indicated in the description frame ("Description") defined further below. These genes in the Description were identified when the designated ORF was compared against a comprehensive non-redundant protein database. Specifically, the sixth column represents the Blast Score ("Score") for the match (a higher score is a better match), and the seventh column represents the

15

20

25

probability ("Probability") for the match (the probability that such a match can have occurred by chance; the lower the value, the more likely the match is valid). If a BLASTP2 score of less than 100 was obtained, no value is reported in the table. The remaining fields below the columns contain additional information relating to the potential function of the sequence based on the BLASTP2 analysis. Where a match was discovered, the field "Protein name" list the protein's name identified from the match. In addition, one skilled in the art would be able to identify the match and elucidate its function using the "Locus name" and where available the accession number, "Acc#" from the database. Lastly, one skilled in the art would appreciate the "Description" field to further describe the potential function of the protein based on this analysis. This information allows one of ordinary skill in the art to determine a potential use for each identified coding sequence and, as a result, allows to use the polypeptides of the present invention for commercial and industrial purposes.

Using the information provided in SEQ ID NO: 1 - SEQ ID NO: 5222, SEQ ID NO: 5223 - SEQ ID NO: 10444 and in Table 2 together with routine cloning and sequencing methods, one of ordinary skill in the art will be able to clone and sequence all the nucleic acid fragments of interest including open reading frames (ORFs) encoding a large variety of proteins of *B. fragilis*.

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility to generate polypeptides. The nucleic acid of the invention exemplified in SEQ ID NO: 1 - SEQ ID NO: 5222 and in Table 2 or fragments of said nucleic acid encoding active portions of *B. fragilis* polypeptides can be cloned into suitable vectors or used to isolate nucleic acid. The isolated nucleic acid is combined with suitable DNA linkers and cloned into a suitable vector.

The function of a specific gene or operon can be ascertained by expression in a bacterial strain under conditions where the activity of the gene product(s) specified by the gene or operon in question can be specifically measured. Alternatively, a gene product may be produced in large quantities in an expressing strain for use as an antigen, an

15

20

25

industrial reagent, for structural studies, etc. This expression can be accomplished in a mutant strain which lacks the activity of the gene to be tested, or in a strain that does not produce the same gene product(s). This includes, but is not limited to, Eucaryotic species such as the yeast *Saccharomyces cerevisiae*, *Methanobacterium* strains or other Archaea, and Eubacteria such as *E. coli*, *B. Subtilis*, *S. Aureus*, *S. Pneumonia* or *Pseudomonas putida*. In some cases the expression host will utilize the natural *B. fragilis* promoter whereas in others, it will be necessary to drive the gene with a promoter sequence derived from the expressing organism (e.g., an *E. coli* beta-galactosidase promoter for expression in *E. coli*).

To express a gene product using the natural *B. fragilis* promoter, a procedure such as the following can be used. A restriction fragment containing the gene of interest, together with its associated natural promoter element and regulatory sequences (identified using the DNA sequence data) is cloned into an appropriate recombinant plasmid containing an origin of replication that functions in the host organism and an appropriate selectable marker. This can be accomplished by a number of procedures known to those skilled in the art. It is most preferably done by cutting the plasmid and the fragment to be cloned with the same restriction enzyme to produce compatible ends that can be ligated to join the two pieces together. The recombinant plasmid is introduced into the host organism by, for example, electroporation and cells containing the recombinant plasmid are identified by selection for the marker on the plasmid. Expression of the desired gene product is detected using an assay specific for that gene product.

In the case of a gene that requires a different promoter, the body of the gene (coding sequence) is specifically excised and cloned into an appropriate expression plasmid. This subcloning can be done by several methods, but is most easily accomplished by PCR amplification of a specific fragment and ligation into an expression plasmid after treating the PCR product with a restriction enzyme or exonuclease to create suitable ends for cloning.

10

15

20

25

A suitable host cell for expression of a gene can be any procaryotic or eucaryotic cell. Suitable methods for transforming host cells can be found in Sambrook et al.

(Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press (1989)), and other laboratory textbooks.

For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding an *B. fragilis* polypeptide can be cultured under appropriate conditions to allow expression of the polypeptide to occur. Suitable media for cell culture are well known in the art. Polypeptides of the invention can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such polypeptides. Additionally, in many situations, polypeptides can be produced by chemical cleavage of a native protein (e.g., tryptic digestion) and the cleavage products can then be purified by standard techniques.

In the case of membrane bound proteins, these can be isolated from a host cell by contacting a membrane-associated protein fraction with a detergent forming a solubilized complex, where the membrane-associated protein is no longer entirely embedded in the membrane fraction and is solubilized at least to an extent which allows it to be chromatographically isolated from the membrane fraction. Chromatographic techniques which can be used in the final purification step are known in the art and include hydrophobic interaction, lectin affinity, ion exchange, dye affinity and immunoaffinity.

One strategy to maximize recombinant *B. fragilis* peptide expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another strategy would be to alter the nucleic acid encoding an *B. fragilis* peptide to be inserted into an expression vector so that the individual codons for each amino acid would be those preferentially utilized in highly expressed *E. coli* proteins (Wada et al., (1992) *Nuc.* 

15

20

25

Acids Res. 20:2111-2118). Such alteration of nucleic acids of the invention can be carried out by standard DNA synthesis techniques.

The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See, e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

The present invention provides a library of *B. fragilis* -derived nucleic acid sequences. The libraries provide probes, primers, and markers which can be used as markers in epidemiological studies. The present invention also provides a library of *B. fragilis* -derived nucleic acid sequences which comprise or encode targets for therapeutic drugs.

Nucleic acids comprising any of the sequences disclosed herein or sub-sequences thereof can be prepared by standard methods using the nucleic acid sequence information provided in SEQ ID NO: 1 - SEQ ID NO: 5222. For example, DNA can be chemically synthesized using, e.g., the phosphoramidite solid support method of Matteucci *et al.*, 1981, *J. Am. Chem. Soc.* 103:3185, the method of Yoo *et al.*, 1989, *J. Biol. Chem.* 764:17078, or other well known methods. This can be done by sequentially linking a series of oligonucleotide cassettes comprising pairs of synthetic oligonucleotides, as described below.

Of course, due to the degeneracy of the genetic code, many different nucleotide sequences can encode polypeptides having the amino acid sequences defined by SEQ ID NO: 5223 - SEQ ID NO: 10444 or sub-sequences thereof. The codons can be selected for optimal expression in prokaryotic or eukaryotic systems. Such degenerate variants are also encompassed by this invention.

Insertion of nucleic acids (typically DNAs) encoding the polypeptides of the invention into a vector is easily accomplished when the termini of both the DNAs and the

15

20

25

vector comprise compatible restriction sites. If this cannot be done, it may be necessary to modify the termini of the DNAs and/or vector by digesting back single-stranded DNA overhangs generated by restriction endonuclease cleavage to produce blunt ends, or to achieve the same result by filling in the single-stranded termini with an appropriate DNA polymerase.

Alternatively, any site desired may be produced, e.g., by ligating nucleotide sequences (linkers) onto the termini. Such linkers may comprise specific oligonucleotide sequences that define desired restriction sites. Restriction sites can also be generated by the use of the polymerase chain reaction (PCR). See, e.g., Saiki *et al.*, 1988, *Science* 239:48. The cleaved vector and the DNA fragments may also be modified if required by homopolymeric tailing.

The nucleic acids of the invention may be isolated directly from cells. Alternatively, the polymerase chain reaction (PCR) method can be used to produce the nucleic acids of the invention, using either chemically synthesized strands or genomic material as templates. Primers used for PCR can be synthesized using the sequence information provided herein and can further be designed to introduce appropriate new restriction sites, if desirable, to facilitate incorporation into a given vector for recombinant expression.

The nucleic acids of the present invention may be flanked by natural *B. fragilis* regulatory sequences, or may be associated with heterologous sequences, including promoters, enhancers, response elements, signal sequences, polyadenylation sequences, introns, 5'- and 3'- noncoding regions, and the like. The nucleic acids may also be modified by many means known in the art. Non-limiting examples of such modifications include methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoroamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.). Nucleic acids may contain one or more additional covalently

20

25

linked moieties, such as, for example, proteins (e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), intercalators (e.g., acridine, psoralen, etc.), chelators (e.g., metals, radioactive metals, iron, oxidative metals, etc.), and alkylators. PNAs are also included. The nucleic acid may be derivatized by formation of a methyl or ethyl phosphotriester or an alkyl phosphoramidate linkage. Furthermore, the nucleic acid sequences of the present invention may also be modified with a label capable of providing a detectable signal, either directly or indirectly. Exemplary labels include radioisotopes, fluorescent molecules, biotin, and the like.

The invention also provides nucleic acid vectors comprising the disclosed *B*. fragilis -derived sequences or derivatives or fragments thereof. A large number of vectors, including plasmid and bacterial vectors, have been described for replication and/or expression in a variety of eukaryotic and prokaryotic hosts, and may be used for cloning or protein expression.

The encoded *B. fragilis* polypeptides may be expressed by using many known vectors, such as pUC plasmids, pET plasmids (Novagen, Inc., Madison, WI), or pRSET or pREP (Invitrogen, San Diego, CA), and many appropriate host cells, using methods disclosed or cited herein or otherwise known to those skilled in the relevant art. The particular choice of vector/host is not critical to the practice of the invention.

Recombinant cloning vectors will often include one or more replication systems for cloning or expression, one or more markers for selection in the host, e.g. antibiotic resistance, and one or more expression cassettes. The inserted *B. fragilis* coding sequences may be synthesized by standard methods, isolated from natural sources, or prepared as hybrids, etc. Ligation of the *B. fragilis* coding sequences to transcriptional regulatory elements and/or to other amino acid coding sequences may be achieved by known methods. Suitable host cells may be transformed/transfected/infected as appropriate by any suitable method including electroporation, CaCl<sub>2</sub> mediated DNA uptake, bacterial infection, microinjection, microprojectile, or other established methods.

20

25

Appropriate host cells include bacteria, archebacteria, fungi, especially yeast, and plant and animal cells, especially mammalian cells. Of particular interest are *B. fragilis*, *E. coli, B. Subtilis, Saccharomyces cerevisiae, Saccharomyces carlsbergensis, Schizosaccharomyces pombi, SF9* cells, C129 cells, 293 cells, *Neurospora*, and CHO cells, COS cells, HeLa cells, and immortalized mammalian myeloid and lymphoid cell lines. Preferred replication systems include M13, ColE1, SV40, baculovirus, lambda, adenovirus, and the like. A large number of transcription initiation and termination regulatory regions have been isolated and shown to be effective in the transcription and translation of heterologous proteins in the various hosts. Examples of these regions, methods of isolation, manner of manipulation, etc. are known in the art. Under appropriate expression conditions, host cells can be used as a source of recombinantly produced *B. fragilis* -derived peptides and polypeptides.

Advantageously, vectors may also include a transcription regulatory element (i.e., a promoter) operably linked to the *B. fragilis* portion. The promoter may optionally contain operator portions and/or ribosome binding sites. Non-limiting examples of bacterial promoters compatible with E. coli include: b-lactamase (penicillinase) promoter; lactose promoter; tryptophan (trp) promoter; araBAD (arabinose) operon promoter; lambda-derived P<sub>1</sub> promoter and N gene ribosome binding site; and the hybrid tac promoter derived from sequences of the trp and lac UV5 promoters. Non-limiting examples of yeast promoters include 3-phosphoglycerate kinase promoter, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) promoter, galactokinase (GAL1) promoter, galactoepimerase promoter, and alcohol dehydrogenase (ADH) promoter. Suitable promoters for mammalian cells include without limitation viral promoters such as that from Simian Virus 40 (SV40), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences, polyA addition sequences and enhancer sequences to increase expression. Sequences which cause amplification of the gene may also be desirable. Furthermore, sequences that facilitate secretion of the recombinant product from cells, including, but

15

20

not limited to, bacteria, yeast, and animal cells, such as secretory signal sequences and/or prohormone pro region sequences, may also be included. These sequences are well described in the art.

Nucleic acids encoding wild-type or variant *B. fragilis* -derived polypeptides may also be introduced into cells by recombination events. For example, such a sequence can be introduced into a cell, and thereby effect homologous recombination at the site of an endogenous gene or a sequence with substantial identity to the gene. Other recombination-based methods such as nonhomologous recombinations or deletion of endogenous genes by homologous recombination may also be used.

The nucleic acids of the present invention find use as templates for the recombinant production of *B. fragilis* -derived peptides or polypeptides.

## IDENTIFICATION AND USE OF B. FRAGILIS NUCLEIC ACID SEQUENCES

The disclosed *B. fragilis* polypeptide and nucleic acid sequences, or other sequences that are contained within ORFs, including complete protein-coding sequences, of which any of the disclosed *B. fragilis* -specific sequences forms a part, are useful as target components for diagnosis and/or treatment of *B. fragilis* - caused infection

It will be understood that the sequence of an entire protein-coding sequence of which each disclosed nucleic acid sequence forms a part can be isolated and identified based on each disclosed sequence. This can be achieved, for example, by using an isolated nucleic acid encoding the disclosed sequence, or fragments thereof, to prime a sequencing reaction with genomic *B. fragilis* DNA as template; this is followed by sequencing the amplified product. The isolated nucleic acid encoding the disclosed sequence, or fragments thereof, can also be hybridized to *B. fragilis* genomic libraries to identify clones containing additional complete segments of the protein-coding sequence of which the shorter sequence forms a part. Then, the entire protein-coding sequence, or fragments thereof, or nucleic acids encoding all or part of the sequence, or sequence-

conservative or function-conservative variants thereof, may be employed in practicing the present invention.

Preferred sequences are those that are useful in diagnostic and/or therapeutic applications. Diagnostic applications include without limitation nucleic-acid-based and antibody-based methods for detecting bacterial infection. Therapeutic applications include without limitation vaccines, passive immunotherapy, and drug treatments directed against gene products that are both unique to bacteria and essential for growth and/or replication of bacteria.

## IDENTIFICATION OF NUCLEIC ACIDS ENCODING VACCINE COMPONENTS AND TARGETS FOR AGENTS EFFECTIVE AGAINST B. FRAGILIS

The disclosed *B. fragilis* genome sequence includes segments that direct the synthesis of ribonucleic acids and polypeptides, as well as origins of replication, promoters, other types of regulatory sequences, and intergenic nucleic acids. The invention encompasses nucleic acids encoding immunogenic components of vaccines and targets for agents effective against *B. fragilis*. Identification of said immunogenic components involved in the determination of the function of the disclosed sequences, which can be achieved using a variety of approaches. Non-limiting examples of these approaches are described briefly below.

20

10

15

### HOMOLOGY TO KNOWN SEQUENCES:

Computer-assisted comparison of the disclosed *B. fragilis* sequences with previously reported sequences present in publicly available databases is useful for identifying functional *B. fragilis* nucleic acid and polypeptide sequences. It will be understood that protein-coding sequences, for example, may be compared as a whole, and that a high degree of sequence homology between two proteins (such as, for example, >80-90%) at the amino acid level indicates that the two proteins also possess some degree of functional homology, such as, for example, among enzymes involved in

15

20

25

metabolism, DNA synthesis, or cell wall synthesis, and proteins involved in transport, cell division, etc. In addition, many structural features of particular protein classes have been identified and correlate with specific consensus sequences, such as, for example, binding domains for nucleotides, DNA, metal ions, and other small molecules; sites for covalent modifications such as phosphorylation, acylation, and the like; sites of protein:protein interactions, etc. These consensus sequences may be quite short and thus may represent only a fraction of the entire protein-coding sequence. Identification of such a feature in an *B. fragilis* sequence is therefore useful in determining the function of the encoded protein and identifying useful targets of antibacterial drugs.

Of particular relevance to the present invention are structural features that are common to secretory, transmembrane, and surface proteins, including secretion signal peptides and hydrophobic transmembrane domains. *B. fragilis* proteins identified as containing putative signal sequences and/or transmembrane domains are useful as immunogenic components of vaccines.

Targets for therapeutic drugs according to the invention include, but are not limited to, polypeptides of the invention, whether unique to *B. fragilis* or not, that are essential for growth and/or viability of *B. fragilis* under at least one growth condition. Polypeptides essential for growth and/or viability can be determined by examining the effect of deleting and/or disrupting the genes, i.e., by so-called gene "knockout".

Alternatively, genetic footprinting can be used (Smith *et al.*, 1995, *Proc. Natl. Acad. Sci. USA* 92:5479-6433; Published International Application WO 94/26933; U.S. Patent No. 5,612,180). Still other methods for assessing essentiality includes the ability to isolate conditional lethal mutations in the specific gene (e.g., temperature sensitive mutations). Other useful targets for therapeutic drugs, which include polypeptides that are not essential for growth or viability *per se* but lead to loss of viability of the cell, can be used to target therapeutic agents to cells.

15

20

25

### STRAIN-SPECIFIC SEQUENCES:

Because of the evolutionary relationship between different *B. fragilis* strains, it is believed that the presently disclosed *B. fragilis* sequences are useful for identifying, and/or discriminating between, previously known and new *B. fragilis* strains. It is believed that other *B. fragilis* strains will exhibit at least about 70% sequence homology with the presently disclosed sequence. Systematic and routine analyses of DNA sequences derived from samples containing *B. fragilis* strains, and comparison with the present sequence allows for the identification of sequences that can be used to discriminate between strains, as well as those that are common to all *B. fragilis* strains. In one embodiment, the invention provides nucleic acids, including probes, and peptide and polypeptide sequences that discriminate between different strains of *B. fragilis*.

In another embodiment, the invention provides nucleic acids, including probes, and peptide and polypeptide sequences that are common to all *B. fragilis* strains but are not found in other bacterial species.

react with antibodies that selectively recognize one or more *B. fragilis* strains.

Strain-specific components can also be identified functionally by their ability to elicit or

#### B. FRAGILIS POLYPEPTIDES

This invention encompasses isolated *B. fragilis* polypeptides encoded by the disclosed *B. fragilis* genomic sequences, including the polypeptides of the invention contained in the Sequence Listing. Polypeptides of the invention are preferably at least about 5 amino acid residues in length. Using the DNA sequence information provided herein, the amino acid sequences of the polypeptides encompassed by the invention can be deduced using methods well-known in the art. It will be understood that the sequence of an entire nucleic acid encoding an *B. fragilis* polypeptide can be isolated and identified based on an ORF that encodes only a fragment of the cognate protein-coding region. This can be achieved, for example, by using the isolated nucleic acid encoding

15

20

25

the ORF, or fragments thereof, to prime a polymerase chain reaction with genomic *B*. *fragilis* DNA as template; this is followed by sequencing the amplified product.

The polypeptides of the present invention, including function-conservative variants of the disclosed ORFs, may be isolated from wild-type or mutant *B. fragilis* cells, or from heterologous organisms or cells (including, but not limited to, bacteria, fungi, insect, plant, and mammalian cells) including *B. fragilis* into which an *B. fragilis* derived protein-coding sequence has been introduced and expressed. Furthermore, the polypeptides may be part of recombinant fusion proteins.

B. fragilis polypeptides of the invention can be chemically synthesized using commercially automated procedures such as those referenced herein, including, without limitation, exclusive solid phase synthesis, partial solid phase methods, fragment condensation or classical solution synthesis. The polypeptides are preferably prepared by solid phase peptide synthesis as described by Merrifield, 1963, J. Am. Chem. Soc. 85:2149. The synthesis is carried out with amino acids that are protected at the alphamino terminus. Trifunctional amino acids with labile side-chains are also protected with suitable groups to prevent undesired chemical reactions from occurring during the assembly of the polypeptides. The alpha-amino protecting group is selectively removed to allow subsequent reaction to take place at the amino-terminus. The conditions for the removal of the alpha-amino protecting group do not remove the side-chain protecting groups.

Methods for polypeptide purification are well-known in the art, including, without limitation, preparative disc-gel electrophoresis, isoelectric focusing, HPLC, reversed-phase HPLC, gel filtration, ion exchange and partition chromatography, and countercurrent distribution. For some purposes, it is preferable to produce the polypeptide in a recombinant system in which the *B. fragilis* protein contains an additional sequence tag that facilitates purification, such as, but not limited to, a polyhistidine sequence. The polypeptide can then be purified from a crude lysate of the host cell by chromatography on an appropriate solid-phase matrix. Alternatively,

15

20

25

antibodies produced against an *B. fragilis* protein or against peptides derived therefrom can be used as purification reagents. Other purification methods are possible.

The present invention also encompasses derivatives and homologues of *B*. *fragilis* -encoded polypeptides. For some purposes, nucleic acid sequences encoding the peptides may be altered by substitutions, additions, or deletions that provide for functionally equivalent molecules, i.e., function-conservative variants. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of similar properties, such as, for example, positively charged amino acids (arginine, lysine, and histidine); negatively charged amino acids (aspartate and glutamate); polar neutral amino acids; and non-polar amino acids.

The isolated polypeptides may be modified by, for example, phosphorylation, sulfation, acylation, or other protein modifications. They may also be modified with a label capable of providing a detectable signal, either directly or indirectly, including, but not limited to, radioisotopes and fluorescent compounds.

To identify *B. fragilis* -derived polypeptides for use in the present invention, essentially the complete genomic sequence of a virulent, methicillin-resistant isolate of *Bacteroides fragilis* isolate was analyzed. While, in very rare instances, a nucleic acid sequencing error may be revealed, resolving a rare sequencing error is well within the art, and such an occurrence will not prevent one skilled in the art from practicing the invention.

Also encompassed are any *B. fragilis* polypeptide sequences that are contained within the open reading frames (ORFs), including complete protein-coding sequences, of which any of SEQ ID NO: 1 - SEQ ID NO: 5222 forms a part. Table 2, which is appended herewith and which forms part of the present specification, provides a putative identification of the particular function of a polypeptide which is encoded by each ORF, based on the homology match (determined by the BLAST algorithm) of the predicted polypeptide with known proteins encoded by ORFs in other organisms. As a result, one

15

20

25

skilled in the art can use the polypeptides of the present invention for commercial and industrial purposes consistent with the type of putative identification of the polypeptide.

The present invention provides a library of *B. fragilis* -derived polypeptide sequences, and a corresponding library of nucleic acid sequences encoding the polypeptides, wherein the polypeptides themselves, or polypeptides contained within ORFs of which they form a part, comprise sequences that are contemplated for use as components of vaccines. Non-limiting examples of such sequences are listed by SEQ ID NO in Table 2, which is appended herewith and which forms part of the present specification.

The present invention also provides a library of *B. fragilis* -derived polypeptide sequences, and a corresponding library of nucleic acid sequences encoding the polypeptides, wherein the polypeptides themselves, or polypeptides contained within ORFs of which they form a part, comprise sequences lacking homology to any known prokaryotic or eukaryotic sequences. Such libraries provide probes, primers, and markers which can be used to diagnose *B. fragilis* infection, including use as markers in epidemiological studies. Non-limiting examples of such sequences are listed by SEQ ID NO in Table 2, which is appended hereto and part hereof.

The present invention also provides a library of *B. fragilis* -derived polypeptide sequences, and a corresponding library of nucleic acid sequences encoding the polypeptides, wherein the polypeptides themselves, or polypeptides contained within ORFs of which they form a part, comprise targets for therapeutic drugs.

## SPECIFIC EXAMPLE: DETERMINATION OF *BACTEROIDES* PROTEIN ANTIGENS FOR ANTIBODY AND VACCINE DEVELOPMENT

The selection of Bacteroides protein antigens for vaccine development can be derived from the nucleic acids encoding *B. fragilis* polypeptides. First, the ORF's can be analyzed for homology to other known exported or membrane proteins and analyzed using the discriminant analysis described by Klein, et al. (Klein, P., Kanehsia, M., and

15

20

DeLisi, C. (1985) *Biochimica et Biophysica Acta* 815, 468-476) for predicting exported and membrane proteins.

Homology searches can be performed using the BLAST algorithm contained in the Wisconsin Sequence Analysis Package (Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) to compare each predicted ORF amino acid sequence with all sequences found in the current GenBank, SWISS-PROT and PIR databases. BLAST searches for local alignments between the ORF and the databank sequences and reports a probability score which indicates the probability of finding this sequence by chance in the database. ORF's with significant homology (e.g. probabilities lower than 1x10 that the homology is only due to random chance) to membrane or exported proteins represent protein antigens for vaccine development. Possible functions can be provided to *B. fragilis* genes based on sequence homology to genes cloned in other organisms.

Discriminant analysis (Klein, et al. supra) can be used to examine the ORF amino acid sequences. This algorithm uses the intrinsic information contained in the ORF amino acid sequence and compares it to information derived from the properties of known membrane and exported proteins. This comparison predicts which proteins will be exported, membrane associated or cytoplasmic. ORF amino acid sequences identified as exported or membrane associated by this algorithm are likely protein antigens for vaccine development.

## PRODUCTION OF FRAGMENTS AND ANALOGS OF *B. FRAGILIS* NUCLEIC ACIDS AND POLYPEPTIDES

Based on the discovery of the *B. fragilis* gene products of the invention provided in the Sequence Listing, one skilled in the art can alter the disclosed structure of *B. fragilis* genes, e.g., by producing fragments or analogs, and test the newly produced structures for activity. Examples of techniques known to those skilled in the relevant art which allow the production and testing of fragments and analogs are discussed below.

10

15

20

25

These, or analogous methods can be used to make and screen libraries of polypeptides, e.g., libraries of random peptides or libraries of fragments or analogs of cellular proteins for the ability to bind *B. fragilis* polypeptides. Such screens are useful for the identification of inhibitors of *B. fragilis*.

#### GENERATION OF FRAGMENTS

Fragments of a protein can be produced in several ways, e.g., recombinantly, by proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of a polypeptide can be generated by removing one or more nucleotides from one end (for a terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes the polypeptide. Expression of the mutagenized DNA produces polypeptide fragments. Digestion with "end-nibbling" endonucleases can thus generate DNAs which encode an array of fragments. DNAs which encode fragments of a protein can also be generated by random shearing, restriction digestion or a combination of the above-discussed methods.

Fragments can also be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, peptides of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or divided into overlapping fragments of a desired length.

#### ALTERATION OF NUCLEIC ACIDS AND POLYPEPTIDES: RANDOM METHODS

Amino acid sequence variants of a protein can be prepared by random mutagenesis of DNA which encodes a protein or a particular domain or region of a protein. Useful methods include PCR mutagenesis and saturation mutagenesis. A library of random amino acid sequence variants can also be generated by the synthesis of a set of degenerate oligonucleotide sequences. (Methods for screening proteins in a library of variants are elsewhere herein).

#### PCR MUTAGENESIS

In PCR mutagenesis, reduced Taq polymerase fidelity is used to introduce random mutations into a cloned fragment of DNA (Leung et al., 1989, *Technique* 1:11-15). The DNA region to be mutagenized is amplified using the polymerase chain reaction (PCR) under conditions that reduce the fidelity of DNA synthesis by Taq DNA polymerase, e.g., by using a dGTP/dATP ratio of five and adding Mn to the PCR reaction. The pool of amplified DNA fragments are inserted into appropriate cloning vectors to provide random mutant libraries.

10

15

20

#### SATURATION MUTAGENESIS

Saturation mutagenesis allows for the rapid introduction of a large number of single base substitutions into cloned DNA fragments (Mayers et al., 1985, *Science* 229:242). This technique includes generation of mutations, e.g., by chemical treatment or irradiation of single-stranded DNA *in vitro*, and synthesis of a complimentary DNA strand. The mutation frequency can be modulated by modulating the severity of the treatment, and essentially all possible base substitutions can be obtained. Because this procedure does not involve a genetic selection for mutant fragments both neutral substitutions, as well as those that alter function, are obtained. The distribution of point mutations is not biased toward conserved sequence elements.

#### DEGENERATE OLIGONUCLEOTIDES

A library of homologs can also be generated from a set of degenerate oligonucleotide sequences. Chemical synthesis of a degenerate sequences can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The synthesis of degenerate oligonucleotides is known in the art (see for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura et al. (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton,

15

20

25

Amsterdam: Elsevier pp273-289; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) *Science* 249:386-390; Roberts et al. (1992) *PNAS* 89:2429-2433; Devlin et al. (1990) *Science* 249: 404-406; Cwirla et al. (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

# ALTERATION OF NUCLEIC ACIDS AND POLYPEPTIDES: METHODS FOR DIRECTED MUTAGENESIS

Non-random or directed, mutagenesis techniques can be used to provide specific sequences or mutations in specific regions. These techniques can be used to create variants which include, e.g., deletions, insertions, or substitutions, of residues of the known amino acid sequence of a protein. The sites for mutation can be modified individually or in series, e.g., by (1) substituting first with conserved amino acids and then with more radical choices depending upon results achieved, (2) deleting the target residue, or (3) inserting residues of the same or a different class adjacent to the located site, or combinations of options 1-3.

### ALANINE SCANNING MUTAGENESIS

Alanine scanning mutagenesis is a useful method for identification of certain residues or regions of the desired protein that are preferred locations or domains for mutagenesis, Cunningham and Wells (*Science* 244:1081-1085, 1989). In alanine scanning, a residue or group of target residues are identified (e.g., charged residues such as Arg, Asp, His, Lys, and Glu) and replaced by a neutral or negatively charged amino acid (most preferably alanine or polyalanine). Replacement of an amino acid can affect the interaction of the amino acids with the surrounding aqueous environment in or outside the cell. Those domains demonstrating functional sensitivity to the substitutions are then refined by introducing further or other variants at or for the sites of substitution.

25

Thus, while the site for introducing an amino acid sequence variation is predetermined, the nature of the mutation per se need not be predetermined. For example, to optimize the performance of a mutation at a given site, alanine scanning or random mutagenesis may be conducted at the target codon or region and the expressed desired protein subunit variants are screened for the optimal combination of desired activity.

#### OLIGONUCLEOTIDE-MEDIATED MUTAGENESIS

Oligonucleotide-mediated mutagenesis is a useful method for preparing substitution, deletion, and insertion variants of DNA, see, e.g., Adelman et al., (*DNA* 2:183, 1983). Briefly, the desired DNA is altered by hybridizing an oligonucleotide encoding a mutation to a DNA template, where the template is the single-stranded form of a plasmid or bacteriophage containing the unaltered or native DNA sequence of the desired protein. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template that will thus incorporate the oligonucleotide primer, and will code for the selected alteration in the desired protein DNA. Generally, oligonucleotides of at least about 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize properly to the single-stranded DNA template molecule. The oligonucleotides are readily synthesized using techniques known in the art such as that described by Crea et al. (*Proc. Natl. Acad. Sci.* USA, 75: 5765[1978]).

#### CASSETTE MUTAGENESIS

Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells et al. (*Gene*, 34:315[1985]). The starting material is a plasmid (or other vector) which includes the protein subunit DNA to be mutated. The codon(s) in the protein subunit DNA to be mutated are identified. There must be a

unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the desired protein subunit DNA. After the restriction sites have been introduced into the plasmid, the plasmid is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures. The two strands are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 3' and 5' ends that are comparable with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now contains the mutated desired protein subunit DNA sequence.

#### COMBINATORIAL MUTAGENESIS

Combinatorial mutagenesis can also be used to generate mutants (Ladner et al., WO 88/06630). In this method, the amino acid sequences for a group of homologs or other related proteins are aligned, preferably to promote the highest homology possible. All of the amino acids which appear at a given position of the aligned sequences can be selected to create a degenerate set of combinatorial sequences. The variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For example, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential sequences are expressible as individual peptides, or alternatively, as a set of larger fusion proteins containing the set of degenerate sequences.

20

10

15

15

20

25

## OTHER MODIFICATIONS OF *B. FRAGILIS* NUCLEIC ACIDS AND POLYPEPTIDES

It is possible to modify the structure of an *B. fragilis* polypeptide for such purposes as increasing solubility, enhancing stability (e.g., shelf life *ex vivo* and resistance to proteolytic degradation *in vivo*). A modified *B. fragilis* protein or peptide can be produced in which the amino acid sequence has been altered, such as by amino acid substitution, deletion, or addition as described herein.

An *B. fragilis* peptide can also be modified by substitution of cysteine residues preferably with alanine, serine, threonine, leucine or glutamic acid residues to minimize dimerization via disulfide linkages. In addition, amino acid side chains of fragments of the protein of the invention can be chemically modified. Another modification is cyclization of the peptide.

In order to enhance stability and/or reactivity, an *B. fragilis* polypeptide can be modified to incorporate one or more polymorphisms in the amino acid sequence of the protein resulting from any natural allelic variation. Additionally, D-amino acids, non-natural amino acids, or non-amino acid analogs can be substituted or added to produce a modified protein within the scope of this invention. Furthermore, an *B. fragilis* polypeptide can be modified using polyethylene glycol (PEG) according to the method of A. Sehon and co-workers (Wie et al., supra) to produce a protein conjugated with PEG.

In addition, PEG can be added during chemical synthesis of the protein. Other modifications of *B. fragilis* proteins include reduction/alkylation (Tarr, *Methods of Protein Microcharacterization*, J. E. Silver ed., Humana Press, Clifton NJ 155-194 (1986)); acylation (Tarr, supra); chemical coupling to an appropriate carrier (Mishell and Shiigi, eds, *Selected Methods in Cellular Immunology*, WH Freeman, San Francisco, CA (1980), U.S. Patent 4,939,239; or mild formalin treatment (Marsh, (1971) *Int. Arch. of Allergy and Appl. Immunol.*, 41: 199 - 215).

To facilitate purification and potentially increase solubility of an *B. fragilis* protein or peptide, it is possible to add an amino acid fusion moiety to the peptide

backbone. For example, hexa-histidine can be added to the protein for purification by immobilized metal ion affinity chromatography (Hochuli, E. et al., (1988) *Bio/Technology*, 6: 1321 - 1325). In addition, to facilitate isolation of peptides free of irrelevant sequences, specific endoprotease cleavage sites can be introduced between the sequences of the fusion moiety and the peptide.

To potentially aid proper antigen processing of epitopes within an *B. fragilis* polypeptide, canonical protease sensitive sites can be engineered between regions, each comprising at least one epitope via recombinant or synthetic methods. For example, charged amino acid pairs, such as KK or RR, can be introduced between regions within a protein or fragment during recombinant construction thereof. The resulting peptide can be rendered sensitive to cleavage by cathepsin and/or other trypsin-like enzymes which would generate portions of the protein containing one or more epitopes. In addition, such charged amino acid residues can result in an increase in the solubility of the peptide.

#### 15 PRIMARY METHODS FOR SCREENING POLYPEPTIDES AND ANALOGS

Various techniques are known in the art for screening generated mutant gene products. Techniques for screening large gene libraries often include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the genes under conditions in which detection of a desired activity, e.g., in this case, binding to *B. fragilis* polypeptide or an interacting protein, facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the techniques described below is amenable to high through-put analysis for screening large numbers of sequences created, e.g., by random mutagenesis techniques.

25

20

#### TWO HYBRID SYSTEMS

Two hybrid assays such as the system described below (as with the other screening methods described herein), can be used to identify polypeptides, e.g.,

15

20

25

fragments or analogs of a naturally-occurring *B. fragilis* polypeptide, e.g., of cellular proteins, or of randomly generated polypeptides which bind to an *B. fragilis* protein. (The *B. fragilis* domain is used as the bait protein and the library of variants are expressed as prey fusion proteins.) In an analogous fashion, a two hybrid assay (as with the other screening methods described herein), can be used to find polypeptides which bind an *B. fragilis* polypeptide.

### **DISPLAY LIBRARIES**

In one approach to screening assays, the Bacteroides peptides are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind an appropriate receptor protein via the displayed product is detected in a "panning assay". For example, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, a detectably labeled ligand can be used to score for potentially functional peptide homologs. Fluorescently labeled ligands, e.g., receptors, can be used to detect homologs which retain ligand-binding activity. The use of fluorescently labeled ligands, allows cells to be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, to be separated by a fluorescence-activated cell sorter.

A gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at concentrations well over phage per milliliter, a large number of phage can be screened at one time. Second, since each infectious phage displays a gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages, M13, fd.,

15

20

25

and f1, are most often used in phage display libraries. Either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle. Foreign epitopes can be expressed at the NH2-terminal end of pIII and phage bearing such epitopes recovered from a large excess of phage lacking this epitope (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

A common approach uses the maltose receptor of *E. coli* (the outer membrane

protein, LamB) as a peptide fusion partner (Charbit et al. (1986) EMBO 5, 3029-3037). Oligonucleotides have been inserted into plasmids encoding the LamB gene to produce peptides fused into one of the extracellular loops of the protein. These peptides are available for binding to ligands, e.g., to antibodies, and can elicit an immune response when the cells are administered to animals. Other cell surface proteins, e.g., OmpA (Schorr et al. (1991) Vaccines 91, pp. 387-392), PhoE (Agterberg, et al. (1990) Gene 88, 37-45), and PAL (Fuchs et al. (1991) *Bio/Tech* 9, 1369-1372), as well as large bacterial surface structures have served as vehicles for peptide display. Peptides can be fused to pilin, a protein which polymerizes to form the pilus-a conduit for interbacterial exchange of genetic information (Thiry et al. (1989) Appl. Environ. Microbiol. 55, 984-993). Because of its role in interacting with other cells, the pilus provides a useful support for the presentation of peptides to the extracellular environment. Another large surface structure used for peptide display is the bacterial motive organ, the flagellum. Fusion of peptides to the subunit protein flagellin offers a dense array of many peptide copies on the host cells (Kuwajima et al. (1988) Bio/Tech. 6, 1080-1083). Surface proteins of other bacterial species have also served as peptide fusion partners. Examples include the Staphylococcus protein A and the outer membrane IgA protease of Neisseria (Hansson et al. (1992) J. Bacteriol. 174, 4239-4245 and Klauser et al. (1990) EMBO J. 9, 1991-1999).

15

20

25

In the filamentous phage systems and the LamB system described above, the physical link between the peptide and its encoding DNA occurs by the containment of the DNA within a particle (cell or phage) that carries the peptide on its surface. Capturing the peptide captures the particle and the DNA within. An alternative scheme uses the DNA-binding protein LacI to form a link between peptide and DNA (Cull et al. (1992) PNAS USA 89:1865-1869). This system uses a plasmid containing the LacI gene with an oligonucleotide cloning site at its 3'-end. Under the controlled induction by arabinose, a LacI-peptide fusion protein is produced. This fusion retains the natural ability of LacI to bind to a short DNA sequence known as LacO operator (LacO). By installing two copies of LacO on the expression plasmid, the LacI-peptide fusion binds tightly to the plasmid that encoded it. Because the plasmids in each cell contain only a single oligonucleotide sequence and each cell expresses only a single peptide sequence, the peptides become specifically and stablely associated with the DNA sequence that directed its synthesis. The cells of the library are gently lysed and the peptide-DNA complexes are exposed to a matrix of immobilized receptor to recover the complexes containing active peptides. The associated plasmid DNA is then reintroduced into cells for amplification and DNA sequencing to determine the identity of the peptide ligands. As a demonstration of the practical utility of the method, a large random library of dodecapeptides was made and selected on a monoclonal antibody raised against the opioid peptide dynorphin B. A cohort of peptides was recovered, all related by a consensus sequence corresponding to a six-residue portion of dynorphin B. (Cull et al. (1992) Proc. Natl. Acad. Sci. U.S.A. 89-1869)

This scheme, sometimes referred to as peptides-on-plasmids, differs in two important ways from the phage display methods. First, the peptides are attached to the C-terminus of the fusion protein, resulting in the display of the library members as peptides having free carboxy termini. Both of the filamentous phage coat proteins, pIII and pVIII, are anchored to the phage through their C-termini, and the guest peptides are placed into the outward-extending N-terminal domains. In some designs, the phage-

20

25

displayed peptides are presented right at the amino terminus of the fusion protein. (Cwirla, et al. (1990) *Proc. Natl. Acad. Sci. U.S.*A. 87, 6378-6382) A second difference is the set of biological biases affecting the population of peptides actually present in the libraries. The LacI fusion molecules are confined to the cytoplasm of the host cells. The phage coat fusions are exposed briefly to the cytoplasm during translation but are rapidly secreted through the inner membrane into the periplasmic compartment, remaining anchored in the membrane by their C-terminal hydrophobic domains, with the N-termini, containing the peptides, protruding into the periplasm while awaiting assembly into phage particles. The peptides in the LacI and phage libraries may differ significantly as a result of their exposure to different proteolytic activities. The phage coat proteins require transport across the inner membrane and signal peptidase processing as a prelude to incorporation into phage. Certain peptides exert a deleterious effect on these processes and are underrepresented in the libraries (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251). These particular biases are not a factor in the LacI display system.

The number of small peptides available in recombinant random libraries is enormous. Libraries of 10 7-10 9 independent clones are routinely prepared. Libraries as large as 10 11 recombinants have been created, but this size approaches the practical limit for clone libraries. This limitation in library size occurs at the step of transforming the DNA containing randomized segments into the host bacterial cells. To circumvent this limitation, an *in vitro* system based on the display of nascent peptides in polysome complexes has recently been developed. This display library method has the potential of producing libraries 3-6 orders of magnitude larger than the currently available phage/phagemid or plasmid libraries. Furthermore, the construction of the libraries, expression of the peptides, and screening, is done in an entirely cell-free format.

In one application of this method (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251), a molecular DNA library encoding 10 decapeptides was constructed and the library expressed in an *E. coli* S30 *in vitro* coupled transcription/translation system.

Conditions were chosen to stall the ribosomes on the mRNA, causing the accumulation

15

20

25

of a substantial proportion of the RNA in polysomes and yielding complexes containing nascent peptides still linked to their encoding RNA. The polysomes are sufficiently robust to be affinity purified on immobilized receptors in much the same way as the more conventional recombinant peptide display libraries are screened. RNA from the bound complexes is recovered, converted to cDNA, and amplified by PCR to produce a template for the next round of synthesis and screening. The polysome display method can be coupled to the phage display system. Following several rounds of screening, cDNA from the enriched pool of polysomes was cloned into a phagemid vector. This vector serves as both a peptide expression vector, displaying peptides fused to the coat proteins, and as a DNA sequencing vector for peptide identification. By expressing the polysome-derived peptides on phage, one can either continue the affinity selection procedure in this format or assay the peptides on individual clones for binding activity in a phage ELISA, or for binding specificity in a completion phage ELISA (Barret, et al. (1992) *Anal. Biochem* 204,357-364). To identify the sequences of the active peptides one sequences the DNA produced by the phagemid host.

### SECONDARY SCREENING OF POLYPEPTIDES AND ANALOGS

The high through-put assays described above can be followed by secondary screens in order to identify further biological activities which will, e.g., allow one skilled in the art to differentiate agonists from antagonists. The type of a secondary screen used will depend on the desired activity that needs to be tested. For example, an assay can be developed in which the ability to inhibit an interaction between a protein of interest and its respective ligand can be used to identify antagonists from a group of peptide fragments isolated though one of the primary screens described above.

Therefore, methods for generating fragments and analogs and testing them for activity are known in the art. Once the core sequence of interest is identified, it is routine for one skilled in the art to obtain analogs and fragments.

15

20

25

#### PEPTIDE MIMETICS OF B. FRAGILIS POLYPEPTIDES

The invention also provides for reduction of the protein binding domains of the subject *B. fragilis* polypeptides to generate mimetics, e.g. peptide or non-peptide agents. The peptide mimetics are able to disrupt binding of a polypeptide to its counter ligand, e.g., in the case of an *B. fragilis* polypeptide binding to a naturally occurring ligand. The critical residues of a subject *B. fragilis* polypeptide which are involved in molecular recognition of a polypeptide can be determined and used to generate *B. fragilis* -derived peptidomimetics which competitively or noncompetitively inhibit binding of the *B. fragilis* polypeptide with an interacting polypeptide (see, for example, European patent applications EP-412,762A and EP-B31,080A).

For example, scanning mutagenesis can be used to map the amino acid residues of a particular B. fragilis polypeptide involved in binding an interacting polypeptide, peptidomimetic compounds (e.g. diazepine or isoquinoline derivatives) can be generated which mimic those residues in binding to an interacting polypeptide, and which therefore can inhibit binding of an B. fragilis polypeptide to an interacting polypeptide and thereby interfere with the function of B. fragilis polypeptide. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides*: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gama lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) J Med Chem 29:295; and Ewenson et al. in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), b-turn dipeptide cores (Nagai et al. (1985) Tetrahedron Lett 26:647; and Sato et al. (1986) J Chem Soc Perkin Trans 1:1231), and b-

(1986) Biochem Biophys Res Commun 134:71).

aminoalcohols (Gordon et al. (1985) Biochem Biophys Res Commun 126:419; and et al.

## VACCINE FORMULATIONS FOR B. FRAGILIS NUCLEIC ACIDS AND POLYPEPTIDES

This invention also features vaccine compositions for protection against infection by *B. fragilis* or for treatment of *B. fragilis* infection. In one embodiment, the vaccine compositions contain one or more immunogenic components such as a surface protein from *B. fragilis*, or portion thereof, and a pharmaceutically acceptable carrier. Nucleic acids within the scope of the invention are exemplified by the nucleic acids of the invention contained in the Sequence Listing which encode *B. fragilis* surface proteins.

Any nucleic acid encoding an immunogenic *B. fragilis* protein, or portion thereof, which is capable of expression in a cell, can be used in the present invention. These vaccines have therapeutic and prophylactic utilities.

One aspect of the invention provides a vaccine composition for protection against infection by *B. fragilis* which contains at least one immunogenic fragment of an *B. fragilis* protein and a pharmaceutically acceptable carrier. Preferred fragments include peptides of at least about 10 amino acid residues in length, preferably about 10-20 amino acid residues in length, and more preferably about 12-16 amino acid residues in length.

Immunogenic components of the invention can be obtained, for example, by screening polypeptides recombinantly produced from the corresponding fragment of the nucleic acid encoding the full-length *B. fragilis* protein. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry.

In one embodiment, immunogenic components are identified by the ability of the peptide to stimulate T cells. Peptides which stimulate T cells, as determined by, for example, T cell proliferation or cytokine secretion are defined herein as comprising at least one T cell epitope. T cell epitopes are believed to be involved in initiation and perpetuation of the immune response to the protein allergen which is responsible for the clinical symptoms of allergy. These T cell epitopes are thought to trigger early events at

15

20

25

the level of the T helper cell by binding to an appropriate HLA molecule on the surface of an antigen presenting cell, thereby stimulating the T cell subpopulation with the relevant T cell receptor for the epitope. These events lead to T cell proliferation, lymphokine secretion, local inflammatory reactions, recruitment of additional immune cells to the site of antigen/T cell interaction, and activation of the B cell cascade, leading to the production of antibodies. A T cell epitope is the basic element, or smallest unit of recognition by a T cell receptor, where the epitope comprises amino acids essential to receptor recognition (e.g., approximately 6 or 7 amino acid residues). Amino acid sequences which mimic those of the T cell epitopes are within the scope of this invention.

Screening immunogenic components can be accomplished using one or more of several different assays. For example, *in vitro*, peptide T cell stimulatory activity is assayed by contacting a peptide known or suspected of being immunogenic with an antigen presenting cell which presents appropriate MHC molecules in a T cell culture. Presentation of an immunogenic *B. fragilis* peptide in association with appropriate MHC molecules to T cells in conjunction with the necessary co-stimulation has the effect of transmitting a signal to the T cell that induces the production of increased levels of cytokines, particularly of interleukin-2 and interleukin-4. The culture supernatant can be obtained and assayed for interleukin-2 or other known cytokines. For example, any one of several conventional assays for interleukin-2 can be employed, such as the assay described in *Proc. Natl. Acad. Sci USA*, 86: 1333 (1989) the pertinent portions of which are incorporated herein by reference. A kit for an assay for the production of interferon is also available from Genzyme Corporation (Cambridge, MA).

Alternatively, a common assay for T cell proliferation entails measuring tritiated thymidine incorporation. The proliferation of T cells can be measured *in vitro* by determining the amount of <sup>3</sup>H-labeled thymidine incorporated into the replicating DNA of cultured cells. Therefore, the rate of DNA synthesis and, in turn, the rate of cell division can be quantified.

15

20

25

Vaccine compositions of the invention containing immunogenic components (e.g., *B. fragilis* polypeptide or fragment thereof or nucleic acid encoding an *B. fragilis* polypeptide or fragment thereof) preferably include a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier that does not cause an allergic reaction or other untoward effect in patients to whom it is administered. Suitable pharmaceutically acceptable carriers include, for example, one or more of water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like, as well as combinations thereof. Pharmaceutically acceptable carriers may further comprise minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives or buffers, which enhance the shelf life or effectiveness of the antibody. For vaccines of the invention containing *B. fragilis* polypeptides, the polypeptide is co-administered with a suitable adjuvant.

It will be apparent to those of skill in the art that the therapeutically effective amount of DNA or protein of this invention will depend, *inter alia*, upon the administration schedule, the unit dose of antibody administered, whether the protein or DNA is administered in combination with other therapeutic agents, the immune status and health of the patient, and the therapeutic activity of the particular protein or DNA.

Vaccine compositions are conventionally administered parenterally, e.g., by injection, either subcutaneously or intramuscularly. Methods for intramuscular immunization are described by Wolff et al. (1990) *Science* 247: 1465-1468 and by Sedegah et al. (1994) *Immunology* 91: 9866-9870. Other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Oral immunization is preferred over parenteral methods for inducing protection against infection by *B. fragilis*. Cain et. al. (1993) *Vaccine* 11: 637-642. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

15

20

25

The vaccine compositions of the invention can include an adjuvant, including, but not limited to aluminum hydroxide; N-acetyl-muramyl--L-threonyl-D-isoglutamine (thr-MDP); N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP); N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphos-phoryloxy)-ethylamine (CGP 19835A, referred to a MTP-PE); RIBI, which contains three components from bacteria; monophosphoryl lipid A; trehalose dimycoloate; cell wall skeleton (MPL + TDM + CWS) in a 2% squalene/Tween 80 emulsion; and cholera toxin. Others which may be used are non-toxic derivatives of cholera toxin, including its B subunit, and/or conjugates or genetically engineered fusions of the *B. fragilis* polypeptide with cholera toxin or its B subunit, procholeragenoid, fungal polysaccharides, including schizophyllan, muramyl dipeptide, muramyl dipeptide derivatives, phorbol esters, labile toxin of *E. coli*, non-*B. fragilis* bacterial lysates, block polymers or saponins.

Other suitable delivery methods include biodegradable microcapsules or immunostimulating complexes (ISCOMs), cochleates, or liposomes, genetically engineered attenuated live vectors such as viruses or bacteria, and recombinant (chimeric) virus-like particles, e.g., bluetongue. The amount of adjuvant employed will depend on the type of adjuvant used. For example, when the mucosal adjuvant is cholera toxin, it is suitably used in an amount of 5 mg to 50 mg, for example 10 mg to 35 mg. When used in the form of microcapsules, the amount used will depend on the amount employed in the matrix of the microcapsule to achieve the desired dosage. The determination of this amount is within the skill of a person of ordinary skill in the art.

Carrier systems in humans may include enteric release capsules protecting the antigen from the acidic environment of the stomach, and including *B. fragilis* polypeptide in an insoluble form as fusion proteins. Suitable carriers for the vaccines of the invention are enteric coated capsules and polylactide-glycolide microspheres. Suitable diluents are 0.2 N NaHCO<sub>3</sub> and/or saline.

15

20

25

Vaccines of the invention can be administered as a primary prophylactic agent in adults or in children, as a secondary prevention, after successful eradication of B. fragilis in an infected host, or as a therapeutic agent in the aim to induce an immune response in a susceptible host to prevent infection by B. fragilis. The vaccines of the invention are administered in amounts readily determined by persons of ordinary skill in the art. Thus, for adults a suitable dosage will be in the range of 10 mg to 10 g, preferably 10 mg to 100 mg. A suitable dosage for adults will also be in the range of 5 mg to 500 mg. Similar dosage ranges will be applicable for children. Those skilled in the art will recognize that the optimal dose may be more or less depending upon the patient's body weight, disease, the route of administration, and other factors. Those skilled in the art will also recognize that appropriate dosage levels can be obtained based on results with known oral vaccines such as, for example, a vaccine based on an E. coli lysate (6 mg dose daily up to total of 540 mg) and with an enterotoxigenic E. coli purified antigen (4 doses of 1 mg) (Schulman et al., J. Urol. 150:917-921 (1993); Boedecker et al., American Gastroenterological Assoc. 999:A-222 (1993)). The number of doses will depend upon the disease, the formulation, and efficacy data from clinical trials. Without intending any limitation as to the course of treatment, the treatment can be administered over 3 to 8 doses for a primary immunization schedule over 1 month (Boedeker, American Gastroenterological Assoc. 888:A-222 (1993)).

In a preferred embodiment, a vaccine composition of the invention can be based on a killed whole *E. coli* preparation with an immunogenic fragment of an *B. fragilis* protein of the invention expressed on its surface or it can be based on an *E. coli* lysate, wherein the killed *E. coli* acts as a carrier or an adjuvant.

It will be apparent to those skilled in the art that some of the vaccine compositions of the invention are useful only for preventing *B. fragilis* infection, some are useful only for treating *B. fragilis* infection, and some are useful for both preventing and treating *B. fragilis* infection. In a preferred embodiment, the vaccine composition of the invention provides protection against *B. fragilis* infection by stimulating humoral

15

20

25

and/or cell-mediated immunity against *B. fragilis*. It should be understood that amelioration of any of the symptoms of *B. fragilis* infection is a desirable clinical goal, including a lessening of the dosage of medication used to treat *B. fragilis* -caused disease, or an increase in the production of antibodies in the serum or mucous of patients.

#### ANTIBODIES REACTIVE WITH B. FRAGILIS POLYPEPTIDES

The invention also includes antibodies specifically reactive with the subject *B. fragilis* polypeptide. Anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, *Antibodies:* A *Laboratory Manual* ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide. Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the subject *B. fragilis* polypeptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies.

In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of the *B. fragilis* polypeptides of the invention, e.g. antigenic determinants of a polypeptide of the invention contained in the Sequence Listing, or a closely related human or non-human mammalian homolog (e.g., 90% homologous, more preferably at least about 95% homologous). In yet a further preferred embodiment of the invention, the anti-*B. fragilis* antibodies do not substantially cross react (i.e., react specifically) with a protein which is for example, less than 80% percent homologous to a sequence of the invention contained in the Sequence Listing. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein which is less than 10 percent, more preferably less than 5 percent, and even more preferably less than 1 percent, of the binding affinity for a protein of the invention

15

20

25

contained in the Sequence Listing. In a most preferred embodiment, there is no cross-reactivity between bacterial and mammalian antigens.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with *B. fragilis* polypeptides. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, F(ab')<sub>2</sub> fragments can be generated by treating antibody with pepsin. The resulting F(ab')<sub>2</sub> fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the invention is further intended to include bispecific and chimeric molecules having an anti-*B. fragilis* portion.

Both monoclonal and polyclonal antibodies (Ab) directed against *B. fragilis* polypeptides or *B. fragilis* polypeptide variants, and antibody fragments such as Fab` and F(ab`)<sub>2</sub>, can be used to block the action of *B. fragilis* polypeptide and allow the study of the role of a particular *B. fragilis* polypeptide of the invention in aberrant or unwanted intracellular signaling, as well as the normal cellular function of the *B. fragilis* and by microinjection of anti-*B. fragilis* polypeptide antibodies of the present invention.

Antibodies which specifically bind *B. fragilis* epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of *B. fragilis* antigens. Anti-*B. fragilis* polypeptide antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate *B. fragilis* levels in tissue or bodily fluid as part of a clinical testing procedure. Likewise, the ability to monitor *B. fragilis* polypeptide levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. The level of an *B. fragilis* polypeptide can be measured in cells found in bodily fluid, such as in urine samples or can be measured in tissue, such as produced by gastric biopsy. Diagnostic assays using anti-*B. fragilis* antibodies can include, for example, immunoassays designed to aid in early diagnosis of *B. fragilis* infections. The

15

20

25

present invention can also be used as a method of detecting antibodies contained in samples from individuals infected by this bacterium using specific *B. fragilis* antigens.

Another application of anti-*B. fragilis* polypeptide antibodies of the invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λgt11, λgt18-23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λgt11 will produce fusion proteins whose amino termini consist of β-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a subject *B. fragilis* polypeptide can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-*B. fragilis* polypeptide antibodies. Phage, scored by this assay, can then be isolated from the infected plate. Thus, the presence of *B. fragilis* gene homologs can be detected and cloned from other species, and alternate isoforms (including splicing variants) can be detected and cloned.

KITS CONTAINING NUCLEIC ACIDS, POLYPEPTIDES OR ANTIBODIES OF THE INVENTION

The nucleic acid, polypeptides and antibodies of the invention can be combined with other reagents and articles to form kits. Kits for diagnostic purposes typically comprise the nucleic acid, polypeptides or antibodies in vials or other suitable vessels. Kits typically comprise other reagents for performing hybridization reactions, polymerase chain reactions (PCR), or for reconstitution of lyophilized components, such as aqueous media, salts, buffers, and the like. Kits may also comprise reagents for sample processing such as detergents, chaotropic salts and the like. Kits may also comprise immobilization means such as particles, supports, wells, dipsticks and the like. Kits may also comprise labeling means such as dyes, developing reagents, radioisotopes, fluorescent agents, luminescent or chemiluminescent agents, enzymes, intercalating agents and the like. With the nucleic acid and amino acid sequence information provided

10

15

20

herein, individuals skilled in art can readily assemble kits to serve their particular purpose. Kits further can include instructions for use.

### **BIO CHIP TECHNOLOGY**

The nucleic acid sequence of the present invention may be used to detect *B*. *fragilis* or other species of *Bacteroides* acid sequence using bio chip technology. Bio chips containing arrays of nucleic acid sequence can also be used to measure expression of genes of *B. fragilis* or other species of *Bacteroides*. For example, to diagnose a patient with a *B. fragilis* or other *Bacteroides* infection, a sample from a human or animal can be used as a probe on a bio chip containing an array of nucleic acid sequence from the present invention. In addition, a sample from a disease state can be compared to a sample from a non-disease state which would help identify a gene that is up-regulated or expressed in the disease state. This would provide valuable insight as to the mechanism by which the disease manifests. Changes in gene expression can also be used to identify critical pathways involved in drug transport or metabolism, and may enable the identification of novel targets involved in virulence or host cell interactions involved in maintenance of an infection. Procedures using such techniques have been described by Brown *et al.*, 1995, *Science* 270: 467-470.

Bio chips can also be used to monitor the genetic changes of potential therapeutic compounds including, deletions, insertions or mismatches. Once the therapeutic is added to the patient, changes to the genetic sequence can be evaluated for its efficacy. In addition, the nucleic acid sequence of the present invention can be used to determine essential genes in cell cycling. As described in Iyer *et al.*, 1999 (*Science*, 283:83-87) genes essential in the cell cycle can be identified using bio chips. Furthermore, the present invention provides nucleic acid sequence which can be used with bio chip technology to understand regulatory networks in bacteria, measure the response to environmental signals or drugs as in drug screening, and study virulence induction.

15

20

25

(Mons *et al.*, 1998, *Nature Biotechnology*, 16: 45-48. Patents teaching this technology include U.S. Patents 5445934, 5744305, and 5800992.

#### DRUG SCREENING ASSAYS USING B. FRAGILIS POLYPEPTIDES

By making available purified and recombinant *B. fragilis* polypeptides, the present invention provides assays which can be used to screen for drugs which are either agonists or antagonists of the normal cellular function, in this case, of the subject *B. fragilis* polypeptides, or of their role in intracellular signaling. Such inhibitors or potentiators may be useful as new therapeutic agents to combat *B. fragilis* infections in humans. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by the person skilled in the art.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with other proteins or change in enzymatic properties of the molecular target. Accordingly, in an exemplary screening assay of the present invention, the compound of interest is contacted with an isolated and purified *B. fragilis* polypeptide.

Screening assays can be constructed *in vitro* with a purified *B. fragilis* polypeptide or fragment thereof, such as an *B. fragilis* polypeptide having enzymatic activity, such that the activity of the polypeptide produces a detectable reaction product. The efficacy of the compound can be assessed by generating dose response curves from

data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. Suitable products include those with distinctive absorption, fluorescence, or chemi-luminescence properties, for example, because detection may be easily automated. A variety of synthetic or naturally occurring compounds can be tested in the assay to identify those which inhibit or potentiate the activity of the *B. fragilis* polypeptide. Some of these active compounds may directly, or with chemical alterations to promote membrane permeability or solubility, also inhibit or potentiate the same activity (e.g., enzymatic activity) in whole, live *B. fragilis* cells.

10

15

20

### **OVEREXPRESSION ASSAYS**

Overexpression assays are based on the premise that overproduction of a protein would lead to a higher level of resistance to compounds that selectively interfere with the function of that protein. Overexpression assays may be used to identify compounds that interfere with the function of virtually any type of protein, including without limitation enzymes, receptors, DNA- or RNA-binding proteins, or any proteins that are directly or indirectly involved in regulating cell growth.

Typically, two bacterial strains are constructed. One contains a single copy of the gene of interest, and a second contains several copies of the same gene. Identification of useful inhibitory compounds of this type of assay is based on a comparison of the activity of a test compound in inhibiting growth and/or viability of the two strains. The method involves constructing a nucleic acid vector that directs high level expression of a particular target nucleic acid. The vectors are then transformed into host cells in single or multiple copies to produce strains that express low to moderate and high levels of protein encoding by the target sequence (strain A and B, respectively). Nucleic acid comprising sequences encoding the target gene can, of course, be directly integrated into the host cell.

15

20

Large numbers of compounds (or crude substances which may contain active compounds) are screened for their effect on the growth of the two strains. Agents which interfere with an unrelated target equally inhibit the growth of both strains. Agents which interfere with the function of the target at high concentration should inhibit the growth of both strains. It should be possible, however, to titrate out the inhibitory effect of the compound in the overexpressing strain. That is, if the compound is affecting the particular target that is being tested, it should be possible to inhibit the growth of strain A at a concentration of the compound that allows strain B to grow.

Alternatively, a bacterial strain is constructed that contains the gene of interest under the control of an inducible promoter. Identification of useful inhibitory agents using this type of assay is based on a comparison of the activity of a test compound in inhibiting growth and/or viability of this strain under both inducing and non-inducing conditions. The method involves constructing a nucleic acid vector that directs high-level expression of a particular target nucleic acid. The vector is then transformed into host cells that are grown under both non-inducing and inducing conditions (conditions A and B, respectively).

Large numbers of compounds (or crude substances which may contain active compounds) are screened for their effect on growth under these two conditions. Agents that interfere with the function of the target should inhibit growth under both conditions. It should be possible, however, to titrate out the inhibitory effect of the compound in the overexpressing strain. That is, if the compound is affecting the particular target that is being tested, it should be possible to inhibit growth under condition A at a concentration that allows the strain to grow under condition B.

### 25 LIGAND-BINDING ASSAYS

Many of the targets according to the invention have functions that have not yet been identified. Ligand-binding assays are useful to identify inhibitor compounds that interfere with the function of a particular target, even when that function is unknown.

10

15

20

25

These assays are designed to detect binding of test compounds to particular targets. The detection may involve direct measurement of binding. Alternatively, indirect indications of binding may involve stabilization of protein structure or disruption of a biological function. Non-limiting examples of useful ligand-binding assays are detailed below.

A useful method for the detection and isolation of binding proteins is the Biomolecular Interaction Assay (BIAcore) system developed by Pharmacia Biosensor and described in the manufacturer's protocol (LKB Pharmacia, Sweden). The BIAcore system uses an affinity purified anti-GST antibody to immobilize GST-fusion proteins onto a sensor chip. The sensor utilizes surface plasmon resonance which is an optical phenomenon that detects changes in refractive indices. In accordance with the practice of the invention, a protein of interest is coated onto a chip and test compounds are passed over the chip. Binding is detected by a change in the refractive index (surface plasmon resonance).

A different type of ligand-binding assay involves scintillation proximity assays (SPA, described in U.S. Patent No. 4,568,649).

Another type of ligand binding assay, also undergoing development, is based on the fact that proteins containing mitochondrial targeting signals are imported into isolated mitochondria *in vitro* (Hurt *et al.*, 1985, *Embo J.* 4:2061-2068; Eilers and Schatz, *Nature*, 1986, 322:228-231). In a mitochondrial import assay, expression vectors are constructed in which nucleic acids encoding particular target proteins are inserted downstream of sequences encoding mitochondrial import signals. The chimeric proteins are synthesized and tested for their ability to be imported into isolated mitochondria in the absence and presence of test compounds. A test compound that binds to the target protein should inhibit its uptake into isolated mitochondria *in vitro*.

Another ligand-binding assay is the yeast two-hybrid system (Fields and Song, 1989, *Nature* 340:245-246). The yeast two-hybrid system takes advantage of the properties of the GAL4 protein of the yeast *Saccharomyces cerevisiae*. The GAL4 protein is a transcriptional activator required for the expression of genes encoding

20

25

enzymes of galactose utilization. This protein consists of two separable and functionally essential domains: an N-terminal domain which binds to specific DNA sequences (UAS<sub>G</sub>); and a C-terminal domain containing acidic regions, which is necessary to activate transcription. The native GAL4 protein, containing both domains, is a potent activator of transcription when yeast are grown on galactose media. The N-terminal domain binds to DNA in a sequence-specific manner but is unable to activate transcription. The C-terminal domain contains the activating regions but cannot activate transcription because it fails to be localized to UAS<sub>G</sub>. In the two-hybrid system, a system of two hybrid proteins containing parts of GAL4: (1) a GAL4 DNA-binding domain fused to a protein 'X' and (2) a GAL4 activation region fused to a protein 'Y'. If X and Y can form a protein-protein complex and reconstitute proximity of the GAL4 domains, transcription of a gene regulated by UAS<sub>G</sub> occurs. Creation of two hybrid proteins, each containing one of the interacting proteins X and Y, allows the activation region of UAS<sub>G</sub> to be brought to its normal site of action.

The binding assay described in Fodor *et al.*, 1991, *Science* 251:767-773, which involves testing the binding affinity of test compounds for a plurality of defined polymers synthesized on a solid substrate, may also be useful.

Compounds which bind to the polypeptides of the invention are potentially useful as antibacterial agents for use in therapeutic compositions.

Pharmaceutical formulations suitable for antibacterial therapy comprise the antibacterial agent in conjunction with one or more biologically acceptable carriers. Suitable biologically acceptable carriers include, but are not limited to, phosphate-buffered saline, saline, deionized water, or the like. Preferred biologically acceptable carriers are physiologically or pharmaceutically acceptable carriers.

The antibacterial compositions include an antibacterial effective amount of active agent. Antibacterial effective amounts are those quantities of the antibacterial agents of the present invention that afford prophylactic protection against bacterial infections or which result in amelioration or cure of an existing bacterial infection. This antibacterial

10

15

effective amount will depend upon the agent, the location and nature of the infection, and the particular host. The amount can be determined by experimentation known in the art, such as by establishing a matrix of dosages and frequencies and comparing a group of experimental units or subjects to each point in the matrix.

The antibacterial active agents or compositions can be formed into dosage unit forms, such as for example, creams, ointments, lotions, powders, liquids, tablets, capsules, suppositories, sprays, aerosols or the like. If the antibacterial composition is formulated into a dosage unit form, the dosage unit form may contain an antibacterial effective amount of active agent. Alternatively, the dosage unit form may include less than such an amount if multiple dosage unit forms or multiple dosages are to be used to administer a total dosage of the active agent. Dosage unit forms can include, in addition, one or more excipient(s), diluent(s), disintegrant(s), lubricant(s), plasticizer(s), colorant(s), dosage vehicle(s), absorption enhancer(s), stabilizer(s), bactericide(s), or the like.

For general information concerning formulations, see, e.g., Gilman et al. (eds.), 1990, Goodman and Gilman's: The Pharmacological Basis of Therapeutics, 8th ed., Pergamon Press; and Remington's Pharmaceutical Sciences, 17th ed., 1990, Mack Publishing Co., Easton, PA; Avis et al. (eds.), 1993, Pharmaceutical Dosage Forms: Parenteral Medications, Dekker, New York; Lieberman et al (eds.), 1990,

20 Pharmaceutical Dosage Forms: Disperse Systems, Dekker, New York.

The antibacterial agents and compositions of the present invention are useful for preventing or treating *B. fragilis* infections. Infection prevention methods incorporate a prophylactically effective amount of an antibacterial agent or composition. A prophylactically effective amount is an amount effective to prevent *B. fragilis* infection and will depend upon the specific bacterial strain, the agent, and the host. These amounts can be determined experimentally by methods known in the art and as described above.

B. fragilis infection treatment methods incorporate a therapeutically effective amount of an antibacterial agent or composition. A therapeutically effective amount is an amount sufficient to ameliorate or eliminate the infection. The prophylactically and/or therapeutically effective amounts can be administered in one administration or over repeated administrations. Therapeutic administration can be followed by prophylactic administration, once the initial bacterial infection has been resolved.

The antibacterial agents and compositions can be administered topically or systemically. Topical application is typically achieved by administration of creams, ointments, lotions, or sprays as described above. Systemic administration includes both oral and parental routes. Parental routes include, without limitation, subcutaneous, intramuscular, intraperitoneal, intravenous, transdermal, inhalation and intranasal administration.

### **EXEMPLIFICATION**

15

20

25

10

### CLONING AND SEQUENCING B. FRAGILIS GENOMIC SEQUENCE

This invention provides nucleotide sequences of the genome of *B. fragilis* which thus comprises a DNA sequence library of *B. fragilis* genomic DNA. The detailed description that follows provides nucleotide sequences of *B. fragilis*, and also describes how the sequences were obtained and how ORFs (Open Reading Frames) and proteincoding sequences can be identified. Also described are methods of using the disclosed *B. fragilis* sequences in methods including diagnostic and therapeutic applications. Furthermore, the library can be used as a database for identification and comparison of medically important sequences in this and other strains of *B. fragilis* as well as other species of *Bacteroides*.

Chromosomal DNA from strain 14062 of *B. fragilis* was isolated after Zymolyase digestion, sodium dodecyl sulfate lysis, potassium acetate precipitation, phenol:chloroform extraction and ethanol precipitation (Soll, D.R., T. Srikantha and S.R.

15

20

Lockhart: Characterizing Developmentally Regulated Genes in *B. fragilis*. In Microbial Genome Methods. K.W. Adolph, editor. CRC Press. New York. p 17-37.). Genomic *B. fragilis* DNA was hydrodynamically sheared in an HPLC and then separated on a standard 1% agarose gel. Fractions corresponding to 2500-3000 bp in length were excised from the gel and purified by the GeneClean procedure (Bio101, Inc.).

The purified DNA fragments were then blunt-ended using T4 DNA polymerase. The healed DNA was then ligated to unique *Bst*XI-linker adapters (5'-GTCTTCACCACGGGG-3' and 5'-GTGGTGAAGAC-3' in 100-1000 fold molar excess). These linkers are complimentary to the *Bst*XI-cut pGTC vector, while the overhang is not self-complimentary. Therefore, the linkers will not concatermerize nor will the cut-vector religate itself easily. The linker-adapted inserts were separated from the unincorporated linkers on a 1% agarose gel and purified using GeneClean. The linker-adapted inserts were then ligated to *Bst*XI-cut vector to construct a "shotgun" sublclone libraries.

Only major modifications to the protocols are highlighted. Briefly, the library was then transformed into DH5 $\alpha$  competent cells (Gibco/BRL, DH5 $\acute{a}$  transformation protocol). It was assessed by plating onto antibiotic plates containing ampicillin and IPTG/Xgal. The plates were incubated overnight at 37 $^{\circ}$ C. Transformants were then used for plating of clones and picking for sequencing. The cultures were grown overnight at 37 $^{\circ}$ C. DNA was purified using a silica bead DNA preparation (Engelstein, 1996) method. In this manner, 25 µg of DNA was obtained per clone.

These purified DNA samples were then sequenced using primarily ABI dyeterminator chemistry. All subsequent steps were based on sequencing by ABI377 automated DNA sequencing methods. The ABI dye terminator sequence reads were run on ABI377 machines and the data was transferred to UNIX machines following lane tracking of the gels. Base calls and quality scores were determined using the program PHRED (Ewing et al., 1998, Genome Res. 8: 175-185; Ewing and Green, 1998, Genome Res. 8: 685-734). Reads were assembled using PHRAP (P. Green, Abstracts of DOE

15

20

25

Human Genome Program Contractor-Grantee Workshop V, Jan. 1996, p.157) with default program parameters and quality scores. The initial assembly was done at 7.8 fold coverage and yielded 223 contigs.

Finishing can follow the initial assembly. Missing mates (sequences from clones that only gave reads from one end of the *Bacteroides* DNA inserted in the plasmid) can be identified and sequenced with ABI technology to allow the identification of additional overlapping contigs.

End-sequencing of randomly picked genomic lambda was also performed. Sequencing on a both sides was done for all lambda sequences. The lambda library backbone helped to verify the integrity of the assembly and allowed closure of some of the physical gaps. Primers for walking off the ends of contigs would be selected using pick\_primer (a GTC program) near the ends of the clones to facilitate gap closure. These walks can be sequenced using the selected clones and primers. These data are then reassembled with PHRAP. Additional sequencing using PCR-generated templates and screened and/or unscreened lambda templates can be done in addition.

To identify *B. fragilis* polypeptides the complete genomic sequence of *B. fragilis* were analyzed essentially as follows: First, all possible stop-to- stop open reading frames (ORFs) greater than 180 nucleotides in all six reading frames were translated into amino acid sequences. Second, the identified ORFs were analyzed for homology to known (archeabacter, prokaryotic and eukaryotic) protein sequences. Third, the coding potential of non-homologous sequences were evaluated with the program GENEMARKTM (Borodovsky and McIninch, 1993, Comp. Chem. 17:123).

## IDENTIFICATION, CLONING AND EXPRESSION OF B. FRAGILIS NUCLEIC ACIDS

Expression and purification of the *B. fragilis* polypeptides of the invention can be performed essentially as outlined below.

20

25

To facilitate the cloning, expression and purification of membrane and secreted proteins from *B. fragilis*, a gene expression system, such as the pET System (Novagen), for cloning and expression of recombinant proteins in *E. coli*, is selected. Also, a DNA sequence encoding a peptide tag, the His-Tag, is fused to the 3' end of DNA sequences of interest in order to facilitate purification of the recombinant protein products. The 3' end is selected for fusion in order to avoid alteration of any 5' terminal signal sequence.

# PCR AMPLIFICATION AND CLONING OF NUCLEIC ACIDS CONTAINING ORF'S ENCODING ENZYMES

Nucleic acids chosen (for example, from the nucleic acids set forth in SEQ ID NO: 1 - SEQ ID NO: 5222 for cloning from the 14062 strain of *B. fragilis* are prepared for amplification cloning by polymerase chain reaction (PCR). Synthetic oligonucleotide primers specific for the 5<sup>1</sup> and 3<sup>1</sup> ends of open reading frames (ORFs) are designed and purchased from GibcoBRL Life Technologies (Gaithersburg, MD, USA). All forward primers (specific for the 5<sup>1</sup> end of the sequence) are designed to include an NcoI cloning site at the extreme 5<sup>1</sup> terminus. These primers are designed to permit initiation of protein translation at a methionine residue followed by a valine residue and the coding sequence for the remainder of the native *B. fragilis* DNA sequence. All reverse primers (specific for the 3<sup>1</sup> end of any *B. fragilis* ORF) include a EcoRI site at the extreme 5<sup>1</sup> terminus to permit cloning of each *B. fragilis* sequence into the reading frame of the pET-28b. The pET-28b vector provides sequence encoding an additional 20 carboxy-terminal amino acids including six histidine residues (at the extreme C-terminus), which comprise the His-Tag.

Genomic DNA prepared from the 14062 strain of *B. fragilis* is used as the source of template DNA for PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). To amplify a DNA sequence containing an *B. fragilis* ORF, genomic DNA (50 nanograms) is introduced into a reaction vial containing 2 mM MgCl<sub>2</sub>, 1 micromolar synthetic oligonucleotide

primers (forward and reverse primers) complementary to and flanking a defined *B. fragilis* ORF, 0.2 mM of each deoxynucleotide triphosphate; dATP, dGTP, dCTP, dTTP and 2.5 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 100 microliters.

Upon completion of thermal cycling reactions, each sample of amplified DNA is washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA). All amplified DNA samples are subjected to digestion with the restriction endonucleases, e.g., NcoI and EcoRI (New England BioLabs, Beverly, MA, USA)(Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). DNA samples are then subjected to electrophoresis on 1.0 % NuSeive (FMC BioProducts, Rockland, ME USA) agarose gels. DNA is visualized by exposure to ethidium bromide and long wave uv irradiation. DNA contained in slices isolated from the agarose gel is purified using the Bio 101 GeneClean Kit protocol (Bio 101 Vista, CA, USA).

15

20

25

10

5

### CLONING OF B. FRAGILIS NUCLEIC ACIDS INTO AN EXPRESSION VECTOR

The pET-28b vector is prepared for cloning by digestion with restriction endonucleases, e.g., NcoI and EcoRI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). The pET-28a vector, which encodes a His-Tag that can be fused to the 5<sup>/</sup> end of an inserted gene, is prepared by digestion with appropriate restriction endonucleases.

Following digestion, DNA inserts are cloned (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994) into the previously digested pET-28b expression vector. Products of the ligation reaction are then used to transform the BL21 strain of *E. coli* (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994) as described below.

10

20

# TRANSFORMATION OF COMPETENT BACTERIA WITH RECOMBINANT PLASMIDS

Competent bacteria, *E coli* strain BL21 or *E. coli* strain BL21(DE3), are transformed with recombinant pET expression plasmids carrying the cloned *B. fragilis* sequences according to standard methods (Current Protocols in Molecular, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). Briefly, 1 microliter of ligation reaction is mixed with 50 microliters of electrocompetent cells and subjected to a high voltage pulse, after which, samples are incubated in 0.45 milliliters SOC medium (0.5% yeast extract, 2.0 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4 and 20, mM glucose) at 37 C with shaking for 1 hour. Samples are then spread on LB agar plates containing 25 microgram/ml kanamycin sulfate for growth overnight. Transformed colonies of BL21 are then picked and analyzed to evaluate cloned inserts as described below.

# 15 IDENTIFICATION OF RECOMBINANT EXPRESSION VECTORS WITH B. FRAGILIS NUCLEIC ACIDS

Individual BL21 clones transformed with recombinant pET-28b *B. fragilis* ORFs are analyzed by PCR amplification of the cloned inserts using the same forward and reverse primers, specific for each *B. fragilis* sequence, that were used in the original PCR amplification cloning reactions. Successful amplification verifies the integration of the *B. fragilis* sequences in the expression vector (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994).

### ISOLATION AND PREPARATION OF NUCLEIC ACIDS FROM

### 25 TRANSFORMANTS

Individual clones of recombinant pET-28b vectors carrying properly cloned *B*. *fragilis* ORFs are picked and incubated in 5 mls of LB broth plus 25 microgram/ml

15

20

25

kanamycin sulfate overnight. The following day plasmid DNA is isolated and purified using the Qiagen plasmid purification protocol (Qiagen Inc., Chatsworth, CA, USA).

### EXPRESSION OF RECOMBINANT B. FRAGILIS SEQUENCES IN E. COLI

The pET vector can be propagated in any *E. coli* K-12 strain e.g. HMS174, HB101, JM109, DH5, etc. for the purpose of cloning or plasmid preparation. Hosts for expression include *E. coli* strains containing a chromosomal copy of the gene for T7 RNA polymerase. These hosts are lysogens of bacteriophage DE3, a lambda derivative that carries the lacI gene, the lacUV5 promoter and the gene for T7 RNA polymerase. T7 RNA polymerase is induced by addition of isopropyl-B-D-thiogalactoside (IPTG), and the T7 RNA polymerase transcribes any target plasmid, such as pET-28b, carrying its gene of interest. Strains used include: BL21(DE3) (Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) Meth. Enzymol. 185, 60-89).

To express recombinant *B. fragilis* sequences, 50 nanograms of plasmid DNA isolated as described above is used to transform competent BL21(DE3) bacteria as described above (provided by Novagen as part of the pET expression system kit). The lacZ gene (beta-galactosidase) is expressed in the pET-System as described for the *B. fragilis* recombinant constructions. Transformed cells are cultured in SOC medium for 1 hour, and the culture is then plated on LB plates containing 25 micrograms/ml kanamycin sulfate. The following day, bacterial colonies are pooled and grown in LB medium containing kanamycin sulfate (25 micrograms/ml) to an optical density at 600 nM of 0.5 to 1.0 O.D. units, at which point, 1 millimolar IPTG was added to the culture for 3 hours to induce gene expression of the *B. fragilis* recombinant DNA constructions.

After induction of gene expression with IPTG, bacteria are pelleted by centrifugation in a Sorvall RC-3B centrifuge at 3500 x g for 15 minutes at  $^{0}$ C. Pellets are resuspended in 50 milliliters of cold 10 mM Tris-HCl, pH 8.0, 0.1 M NaCl and 0.1 mM EDTA (STE buffer). Cells are then centrifuged at 2000 x g for 20 min at  $^{0}$ C. Wet pellets are weighed and frozen at  $^{0}$ C until ready for protein purification.

20

25



A variety of methodologies known in the art can be utilized to purify the isolated proteins. (Current Protocols in Protein Science, John Wiley and Sons, Inc., J. E. Coligan et al., eds., 1995). For example, the frozen cells may be thawed, resupended in buffer and ruptured by several passages through a small volume microfluidizer (Model M-110S, Microfluidics International Corporation, Newton, MA). The resultant homogenate may be centrifuged to yield a clear supernatant (crude extract) and following filtration the crude extract may be fractionated over columns. Fractions may be monitored by absorbance at OD280 nm. and peak fractions may analyzed by SDS-PAGE

The concentrations of purified protein preparations may be quantified spectrophotometrically using absorbance coefficients calculated from amino acid content (Perkins, S.J. 1986 Eur. J. Biochem. 157, 169-180). Protein concentrations are also measured by the method of Bradford, M.M. (1976) Anal. Biochem. 72, 248-254, and Lowry, O.H., Rosebrough, N., Farr, A.L. & Randall, R.J. (1951) J. Biol. Chem. 193, pages 265-275, using bovine serum albumin as a standard.

SDS-polyacrylamide gels of various concentrations may be purchased from BioRad (Hercules, CA, USA), and stained with Coomassie blue. Molecular weight markers may include rabbit skeletal muscle myosin (200 kDa), *E. coli* (-galactosidase (116 kDa), rabbit muscle phosphorylase B (97.4 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), bovine carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa), egg white lysozyme (14.4 kDa) and bovine aprotinin (6.5 kDa).

### **EQUIVALENTS**

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments and methods described herein. The specific embodiments described herein are offered by way of example only, and the invention is to limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

### TABLE 2

ODE Name	NTID AAID	NT AA Score	Probability
ORF Name		139   420   196	1.5e-15
32245287 <u></u> ±2 <u></u> 21	5223		
<u>Protein name</u>		Locus Name	Acc#
hypothetical protein	jhp1211	pir:C71832	C71832
Description			
		NT AA Grana	
ORF Name	NTID AAID	Length Length	Probability
10020167_c1_80	5224	611 1836 706	1.4e-69
Protein name	·	Locus Name	Acc#
glutaminase A	_*	gp:AB029552	AB029552
Description			
Aspergillus oryzae g	taA gene for glut	aminase A, complete cds	•
		NT AA	Decelerated Library
ORF Name	NTID AAID	NT AA Score Length Length	Probability
	NTID AAID 3 5225	— <u> </u>	Probability
		Length Length Score	-
1064765_c1_89		<u>Length</u> <u>Length</u> <u>Score</u> 249 750 324	1.8e-28
1064765_c1_89		Length Length Score  249 750 324  Locus Name	1.8e-28 <u>Acc#</u>
Protein name  alpha-1,6-mannanase  Description	5225	Length Length Score  249 750 324  Locus Name	1.8e-28 <u>Acc#</u> AB024331
Protein name  alpha-1,6-mannanase  Description	5225	Length Length  249  750  324  Locus Name  gp:AB024331  bha-1,6-mannanase, compl	1.8e-28 Acc# AB024331 etecds.
Protein name  alpha-1,6-mannanase  Description	5225	Length Length Score    249   750   324     Locus Name       gp:AB024331	1.8e-28 Acc# AB024331 etecds.
Protein name  alpha-1,6-mannanase  Description  Bacillus circulans a	5225 man6 gene for alp	Length Length  249 750 324  Locus Name  gp:AB024331  Sha-1,6-mannanase, compl	1.8e-28 Acc# AB024331 etecds.
Protein name  alpha-1,6-mannanase  Description  Bacillus circulans a	man6 gene for alp	Length Length  249 750 324  Locus Name  gp:AB024331  bha-1,6-mannanase, compl  NT AA Length Length Score	1.8e-28 Acc# AB024331 etecds.
Protein name  alpha-1,6-mannanase  Description  Bacillus circulans a  ORF Name  10945326_f3_56  Protein name	man6 gene for alp	Length Length  249 750 324  Locus Name  gp:AB024331  Sha-1,6-mannanase, compl  NT AA Length Length  Score  481 1446	Acc# AB024331 etecds.  Probability
Protein name  alpha-1,6-mannanase  Description  Bacillus circulans a  ORF Name	man6 gene for alp	Length Length  249 750 324  Locus Name  gp:AB024331  Sha-1,6-mannanase, compl  NT AA Length Length  Score  481 1446	Acc# AB024331 etecds.  Probability

ORF Name	NTID AAID	NT AA Length Length	Score	Probability
12109430_c2_116	5 5227	788 2367	2343	4.6e-243
Protein name		Locu	s Name	Acc#
immunoreactive 89kD	antigen PG87	gp:AF1	75722	AF175722
Description				
Porphyromonas gingi complete cds.	valis strain W50 i	mmunoreactive 89	kD antiger	iPG87 gene,
		NT AA	Score	Probability
ORF Name	NTID AAID	<u>Length</u> <u>Length</u>	<u>Beere</u>	×
14647327_c1_92	5228	2514	706	2.7e-76
Protein name		Loci	ıs Name	Acc#
glutaminase A		gp:AB0	29552	AB029552
Description				
Aspergillus oryzae	gtaA gene for glut	aminase A, comp	ete cds.	
Abpergrade or just	J J			
Appergration		NT AA		Probability
ORF Name	NTID AAID		Score	Probability
		NT AA	Score	Probability 2.6e-75
ORF Name	NTID AAID	NT AA Length Length	Score	19
ORF Name 19125_c1_98	NTID AAID 7 5229	NT AA Length Length	Score 760 us Name	2.6e-75
ORF Name  19125_c1_98  Protein name	NTID AAID 7 5229	NT AA Length Length 406 1221 Loc	Score 760 us Name	2.6e-75 <u>Acc#</u>
ORF Name  19125_cl_98  Protein name  putative aldose 1-6	NTID AAID  [5229]  Epimerase	NT AA Length Length 406 1221 Loc	Score 760 us Name	2.6e-75 <u>Acc#</u>
ORF Name  19125_cl_98  Protein name  putative aldose 1-e  Description	NTID AAID  [5229]  Epimerase	NT AA Length Length  406 1221  Local gp:SC	Score 760 us Name 4A7	2.6e-75 <u>Acc#</u> AL133423
ORF Name  19125_cl_98  Protein name  putative aldose 1-e  Description	NTID AAID  [5229]  Epimerase	NT AA Length Length 406 1221  Local	Score 760 us Name 4A7	Acc# AL133423  Probability
ORF Name  19125_cl_98  Protein name  putative aldose 1-e  Description  Streptomyces coelic	NTID AAID  7 5229  Epimerase  Color cosmid 4A7.  NTID AAID	NT AA Length Length  406 1221  Local gp:SC	Score    760   18 Name   4A7	2.6e-75 <u>Acc#</u> AL133423
ORF Name  19125_cl_98  Protein name  putative aldose 1-e  Description  Streptomyces coelic	NTID AAID  7 5229  Epimerase  Color cosmid 4A7.  NTID AAID	NT AA Length Length  Loc  gp:SC  NT AA  Length  Loc  1221  Loc  1085  3258	Score 760 as Name 4A7 Score	2.6e-75  Acc#  AL133423  Probability  1.9e-78  Acc#
ORF Name  19125cl98  Protein name  putative aldose 1-e  Description  Streptomyces coelic  ORF Name  19536316c2107	NTID AAID 7 5229 Epimerase Color cosmid 4A7.  NTID AAID 8 5230	NT AA Length Length  406 1221  Loc  gp:SC  NT AA Length Length  1085 3258  Loc	Score 760  IS Name 4A7  Score 755  US Name	2.6e-75  Acc#  AL133423  Probability  1.9e-78

ORF Name NTID AAID Length Length Probability	
22455343_c1_88 9 5231 65 198 55 0.031	
Protein name Locus Name Acc#	
gp:AP000969 AP000969	
Description	
Oryza sativa genomic DNA, chromosome 1, clone:P0011D01.	
NT AA Score Probability	_
ORF Name NTID AAID Length Length Probability	
22457686_c2_112	
Protein name Locus Name Acc#	
Description	
NO-HIT	
NT AA Grave Probability	
ORF Name NTID AAID Length Length Probability	
23.4.9.2.7.8.6	
Protein name Locus Name Acc#	
Description	
ро-ніт	
NIII 7.7	
23866437 <u></u> c3 <u></u> 134 <u></u>	
Protein name Locus Name Acc#	
hypothetical protein SCJ4.42c pir:T37125	
Description	
NT AA Grand Drobability	
ORF Name NTID AAID Length Length Probability	
24645308_f1_2013	
Protein name Locus Name Acc#	
hypothetical protein pir:S76045 S76045	
Description	

	NTID AAID Length Length Score Probability
ORF Name	[6.00-11]
24647811_c2_108	
Protein name	Locus Name Acc#
unknown	gp:U96771 U96771
Description	
Prevotella bryan mannanase genes,	tii putative polygalacturonase,B-1,4-endoglucanase, and complete cds; and unknowngenes.
	NT AA Score Probability
ORF Name	NTID AAID Length Length
26307018_c2_119	15 5237 427 1284 351 5.6e-32
Protein name	Locus Name Acc#
	sp:YXAH_BACSU P42107
Description	
HYPOTHETICAL 46	2 KD PROTEIN IN ASNH-GNTR INTERGENIC REGION
	NT AA Grand Drobability
ORF Name	NTID AAID Length Length Score Probability
29461537c3138	16   5238   158   477
Protein name	Locus Name Acc#
Description	
NO-HIT	
ORF Name	$rac{ ext{NT}}{ ext{NTID}}$ $rac{ ext{AAID}}{ ext{Length}}$ $rac{ ext{Length}}{ ext{Length}}$ $rac{ ext{Score}}{ ext{Probability}}$
3027177 <u></u> .c1 <u></u> 87	
Protein name	Locus Name Acc#
endo-arabinase	gp:D85132 D85132
Description	
Bacillus subtil	is DNA for endo-arabinase, complete cds.

ORF Name NTID AAID Length Length	Probability
ORF Name NTID AAID Length Hength 14142127 c2 109 18 5240 147 444	
Protein name Locus Name	Acc#
Flotern mane	
Description	
NO-HIT	
	Probability
4492168 <u></u> c3 <u></u> 125	2.3e-17
Protein name Locus Name	Acc#
alpha-1,6-mannanase gp:AB024331	AB024331
Description	
Bacillus circulans aman6 gene for alpha-1,6-mannanase, comple	tecds.
27	
$\frac{\text{NT}}{\text{NT}}$ Score	Probability
ORF Name NTID AAID Length Length	Probability
ORF Name         NTID         AAID         Length         Length           47.266.87cl86	
ORF Name NTID AAID Length Length	Probability  Acc#
ORF Name         NTID         AAID         Length         Length           47.266.87cl86	
ORF Name         NTID         AAID         Length         Length           47.266.87cl86	
ORF Name         NTID         AAID         Length         Length           47.266.87c186	Acc#
ORF Name         NTID         AAID         Length         Length           47.266.87cl86	
ORF Name         NTID         AAID         Length         Length           4.726.6.8.7c18.6	Acc# Probability
ORF Name         NTID         AAID         Length         Length           4726687c186	Acc#
ORF Name         NTID         AAID         Length         Length           4.726.6.8.7c186	Acc# Probability

ORF Name	NTID AAID	NT Length	AA Length Score	Probability
12203438_c1_6	22 5244	420 1	.263 1627	3.4e-167
Protein name			Locus Name	Acc#
hypothetical prote	in		pir:JQ1020	JQ1020
Description				
ORF Name	NTID AAID	<u>NT</u> Length	AA Length Score	Probability
26365816 <u></u> £3 <u></u> 4	. 23 5245	857 2	574	
Protein name			Locus Name	Acc#
Description				
NO-HIT	-			
		NT	AA ~	
ORF Name	NTID AAID	Length	Length Score	Probability
32631533_f1_1	24 5246	295	435	2.3e-48
32631533_f1_1  Protein name	.] 24 5246	295	Locus Name	2.3e-48 Acc#
Protein name				
Protein name glucan 1,4-beta-glucosidas			Locus Name	Acc#
Protein name			Locus Name	Acc#
Protein name glucan 1,4-beta-glucosidas			Locus Name	Acc#
Protein name glucan 1,4-beta-glucosidas Description	e,:exo-1,4-beta-g	Lucosidase  NT  Length	Locus Name  pir:JC4825  AA Score	Acc# JC4825
Protein name  glucan 1,4-beta-glucosidas  Description  ORF Name	e,:exo-1,4-beta-g	Lucosidase  NT  Length	Locus Name  pir:JC4825  AA Length Score	Acc# JC4825  Probability
Protein name  glucan 1,4-beta-glucosidas  Description  ORF Name  22275187tll	e,:exo-1,4-beta-g	Lucosidase  NT  Length	Locus Name  pir:JC4825  AA Length Score  13 482	Acc# JC4825  Probability 7.4e-46
Protein name  glucan 1,4-beta-glucosidas  Description  ORF Name  22275187_fl_1  Protein name	e,:exo-1,4-beta-g	Lucosidase  NT  Length	Locus Name  pir:JC4825  AA Length Score  Locus Name	Acc# JC4825  Probability  7.4e-46 Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24432962_f2_1 Protein name	26	5248	91 2	Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
31767067 <u></u> ±1 <u></u> 1	27	5249	294	885		*
Protein name				Locus	s Name	Acc#
Description						
NO-HIT						
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>		*
14257180 <u></u> f2 <u></u> 19	28	5250	493	1482	250	6.5e-36
Protein name				Locus	s Name	Acc#
	•					
				sp:ARS	F_HUMAN	P54793
Description				sp:ARS	F_HUMAN	P54793
Description  ARYLSULFATASE F PR	ECURSOR, (	ASF)	7	sp:ARS	F_HUMAN	P54793
	ECURSOR, (	(ASF)	NT	sp:ARS		
	ECURSOR, (	(ASF) <u>AAID</u>	NT Length		Score	P54793  Probability
ARYLSULFATASE F PR	NTID		Length	AA		
ARYLSULFATASE F PR	NTID	AAID	Length	AA Length	Score	Probability
ARYLSULFATASE F PR ORF Name 24025282c275	NTID	AAID	Length	AA Length 1653	Score	Probability 1.6e-105
ARYLSULFATASE F PR ORF Name 24025282c275	NTID	AAID	Length	AA Length 1653	Score 1045 s Name	Probability  1.6e-105  Acc#

ORF Name NTID AAID Length Length	lity
24486016 ±2 21   30   5252   980   2943   165   1.4e-	09
Protein name Locus Name Ac	cc#
response regulator gp:SPAJ6398 AJ0	006398
Description	
Streptococcus pneumoniae rr09 and hk09 genes; two component system09.	
NT AA Grave Probabi	
ORF Name NTID AAID Length Length Probabi	lity
25584525_f1_10 31 5253 786 2361 212 2.8e-	-14
Protein name Locus Name Ac	cc#
putative secreted protein gp:SCF41 AL1	117387
Description	
Streptomyces coelicolor cosmid F41.	
NT AA	
ORF Name NTID AAID Length Length	ility
Score Propabilities of the second sec	
ORF Name         NTID         AAID         Length         Length         Score         Probabilities           26209530_c3_98	
ORF Name         NTID         AAID         Length         Length         Score         Probabilities           26209530c398	-32
ORF Name         NTID         AAID         Length         Length         Score         Probabilities           26209530_c3_98	-32 cc# 4852
ORF Name         NTID         AAID         Length         Length         Score         Probabilities           26209530_c3_98	-32 cc# 4852
ORF Name  NTID  AAID  Length  Length  26.20.9530c398	-32 cc# 4852 gene,
ORF Name  NTID  AAID  Length  Length  26209530c398	-32 cc# 4852 gene,
ORF Name  NTID  AAID  Length  Length  26209530c398	-32 cc# 4852 gene,
ORF Name  NTID  AAID  Length  Length  Score  Probability  2620.953.0c398	-32 cc# 4852 gene,
ORF Name         NTID         AAID         Length         Length         Score         Probabilities           2620.9530c398	-32 cc# 4852 gene,

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2932812_c2_68	34	5256	509	L <b>5</b> 30	258	1.4e-34
Protein name				Locus	s Name	Acc#
				sp:ARS	E_HUMAN	P51690
Description						
ARYLSULFATASE E PI	RECURSOR, (	ASE)				
			NT	AA	_	D b b l l l l l l
ORF Name	NTID	AAID	Length	Length	Score	Probability
3020203_c2_67	35	5257	423	1272	666	2.3e-65
Protein name	•			Locu	s Name	Acc#
	4			sp:HEX	A_PORGI	P49008
Description						
(BETA-NAHASE)						Ĭ.
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
3949012 <u></u> c2 <u>69</u>	36	5258	487	1464	726	1.0e-71
Protein name				Locu	s Name	Acc#
				sp:MOD	F_ECOLI	P31060
Description						
PROTEIN PHRA)			,		*	-
		1	NT	AA		**
ORF Name		AAID	Length	Length	Score	Probability
	NTID	FIFTID				
47.26588 <u></u> c3 <u></u> 86		5259		1113	1233	1.9e-125
4.7.26.5.88C38.6 Protein name				1113	1233 s Name	1.9e-125 <u>Acc#</u>
	37			1113	s Name	

15
is ones
ĻŊ
2000
24 Hzt
fij
4 2
祌
383
sai tite
13
122
100 City
W 199

		NT	AA Score	Probability
ORF Name	NTID AAI	D Length	Length BCOIC	
5273452_c1_61	38 5260	248 7	783	9.4e-78
Protein name			Locus Name	Acc#
			ap:PMG1_ECOLI	P31217
Description				
(PGAM 1) (BPG-DE	PENDENT PGAM 1)			-
		NT	AA	Drobability
ORF Name	NTID AA		Length Score	Probability
10569587_c1_59	39 5261	768	2307 924	1.1e-92
Protein name			Locus Name	Acc#
melibiase			gp:TEMELA	Y08557
Description				
T.ethanolicus me	IA and lacA genes	•		
T.ethanolicus me	IA and lacA genes		AA	
T.ethanolicus me	IA and lacA genes  NTID AA	NT	AA Length Score	Probability
	NTID AA	NT ID Length	Score	Probability
ORF Name	NTID AA	NT ID Length	Length Score	Probability  Acc#
ORF Name  136.9077_t3_49  Protein name	NTID AA	NT ID Length	Length Score	
ORF Name  136.90.77t349  Protein name  Description	NTID AA	NT ID Length	Length Score	
ORF Name  136.9077_t3_49  Protein name	NTID AA	NT ID Length	Length Score	
ORF Name  1369077_f3_49  Protein name  Description  NO-HIT	NTID AA	NT Length 405	Length  Locus Name  AA Score	
ORF Name  1369077_f3_49  Protein name  Description  NO-HIT  ORF Name	NTID AA 40 5262	NT Length  405  NT Length  Length	Length  Locus Name	Acc#
ORF Name  1369077_f3_49  Protein name  Description  NO-HIT  ORF Name  15631576_c1_74	NTID AA 40 5262	NT Length  405  NT Length  Length	Length  Locus Name  Locus Name  AA Length  Score  AI 1203	Acc# Probability
ORF Name  136.90.77t349  Protein name  Description  NO-HIT  ORF Name  156.315.76c174  Protein name	NTID AA 40   5262  NTID AA 41   5263	NT Length  405  NT Length  Length	Length  Locus Name  Locus Name  AA Length  Locus Name  Locus Name	Acc#  Probability  0.0016  Acc#
ORF Name  136.90.77t349  Protein name  Description  NO-HIT  ORF Name  156.315.76c174  Protein name	NTID AA 40 5262	NT Length  405  NT Length  Length	Length  Locus Name  Locus Name  AA Length  Score  AI 1203	Acc# Probability  0.0016

]
,
]
_
]
]
]
]
) 2

ORF Name	NTID AAID	NT A	- Score	Probability
2054635_f2_34	16 5268	115 348	353	3.4e-32
Protein name			ocus Name	Acc#
hypothetical protein	b0866	pir	:B64825	B64825
Description				·
		NT A	A	
ORF Name	NTID AAID	Length Length	– Score	Probability
20.7240.02c16.3	5269	553 1662	742	2.1e-73
Protein name		Ī	ocus Name	Acc#
probable secreted al	pha-galactosidase	piı	:T36472	T36472
Description				
		NT A	Δ	
ORF Name	NTID AAID	Length Len	- Score	Probability
22519677 <u>f1</u> 3	48 5270	1350	1532	4.0e-157
Protein name		<u> </u>	Locus Name	Acc#
L-fucose permease		gp	:AF137263	AF137263
Description				*
Bacteroides thetaiot gene cluster, and RNA complete cds.	aomicron 30S ribo A polymerase sigma	somal proteir factorSigZ-l	i S16-likepr Like protein	otein, fucose (sigZ) genes,
complete cus.				
ORF Name	NTID AAID		A Score	Probability
22917183_c1_62	49 5271	573 1722	297	9.3e-23
Protein name			Locus Name	Acc#
receptor antigen (Ra	ıgA)	gp	:PGI130872	AJ130872
Description				
Porphyromonas gingiv immunodominant 55kDa	valis W50 receptor antigen.	antigen (rag	g) locus enc	odinga major

NT AA Score Probability
ORF Name NTID AAID Length Length
23679512_c1_61
Protein name Locus Name Acc#
115K outer membrane protein precursor:SusC pir:JC6027 JC6027
protein
Description
NT AA Ggara Probability
ORF Name NTID AAID Length Length Probability
24490677c3119
Protein name Locus Name Acc#
probable sigK protein pir:F70830 F70830
Description
<u>Description</u>
NTID AAID Length Length Probability
ORF Name
24640927c312152 5274 616 1851 116 0.00016
Protein name Locus Name Acc#
unknown gp:U96771 U96771
Description
Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and
mannanase genes, complete cds; and unknowngenes.
NT AA Score Probability
ORF Name NTID AAID Length Length
24803457 <u></u> c1 <u></u> 65
Protein name Locus Name Acc#
Description
Description
NO-HIT

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
25963962 f2_28	54	5276	377	L134	123	0.00063	
Protein name		<u> </u>	L	Locus	Name	Acc#	
11000111				sp:TRH	/ RABIT	P37709	
Description							
TRICHOHYALIN	*						
			NIII	7) 7)			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
26212805_f2_29	55	5277	156	471.	119	2.2e-07	
Protein name				Locus	s Name	Acc#	
				sp:YHB	C_ECOLI	P03843	
Description						·	
HYPOTHETICAL 16.8	8 KD PROTEIN	IN NUSA-	METY INT	ERGENIC I	REGION		1
							ļ
				AA		Dl-abilites	]
ORF Name	NTID	AAID	<u>NT</u> Length		Score	Probability	<u> </u>
	NTID		NT Length	AA		Probability 7.7e-186	
ORF Name	NTID	AAID	NT Length	AA Length 1566	Score		
ORF Name 26364056t231	NTID	AAID	NT Length	<u>AA</u> Length 1566	Score	7.7e-186	
ORF Name 26364056t231	NTID	AAID	NT Length	<u>AA</u> Length 1566	Score 1803 s Name	7.7e-186 <u>Acc#</u>	, , , , , , , , , , , , , , , , , , ,
ORF Name  26364056t231  Protein name	<u>NTID</u> 	<u>AAID</u> 5278	NT Length	<u>AA</u> Length 1566	Score 1803 s Name	7.7e-186 <u>Acc#</u>	
ORF Name  26364056_t2_31  Protein name  Description	<u>NTID</u> 	<u>AAID</u> 5278	NT Length	<u>AA</u> Length 1566	Score  [1803] S Name  4_SYNY3	7.7e-186 <u>Acc#</u> Q55790	
ORF Name  26364056_t2_31  Protein name  Description	<u>NTID</u> 	<u>AAID</u> 5278	NT Length 521	AA Length 1566 Locu sp:Y07	Score 1803 s Name	7.7e-186 <u>Acc#</u>	
ORF Name  26.3640.56f231  Protein name  Description  HYPOTHETICAL 52.	NTID [56  8 KD PROTEIN NTID	AAID 5278 SLR0074	NT Length  NT Length	AA Length  1566  Locu  sp:Y07	Score  [1803] S Name  4_SYNY3	7.7e-186 <u>Acc#</u> Q55790	
ORF Name  26.364.056f231  Protein name  Description  HYPOTHETICAL 52.	NTID [56  8 KD PROTEIN NTID	AAID 5278 SLR0074 AAID	NT Length 521 NT Length	AA Length  Locu  sp:Y07  AA Length	Score  [1803] S Name  4_SYNY3	7.7e-186 <u>Acc#</u> Q55790	
ORF Name  26.364.0.56£231  Protein name  Description  HYPOTHETICAL 52.  ORF Name  2948.0.3.06£351	NTID [56  8 KD PROTEIN NTID	AAID 5278 SLR0074 AAID	NT Length 521 NT Length	AA Length  Locu  sp:Y07  AA Length	Score  1803  S Name  4_SYNY3  Score	7.7e-186 <u>Acc#</u> Q55790  Probability	
ORF Name  26.36.40.56t231  Protein name  Description  HYPOTHETICAL 52.  ORF Name  29.480.306t351  Protein name	NTID [56  8 KD PROTEIN NTID	AAID 5278 SLR0074 AAID	NT Length 521 NT Length	AA Length  Locu  sp:Y07  AA Length	Score  1803  S Name  4_SYNY3  Score	7.7e-186 <u>Acc#</u> Q55790  Probability	

		NT AA Grand Drobability
ORF Name	NTID AAID	Length Length Score Probability
30578126_f1_4	58 5280	446 1341
Protein name		Locus Name Acc#
Description		
NO-HIT		
ORF Name	NTID AAID	NT AA Score Probability Length Length
31663925 <u></u> £3 <u></u> 50	59 5281	89 270 76 0.0077
Protein name		Locus Name Acc#
probable serine p	roteinase	pir:T36552 T36552
Description		
ORF Name	NTID AAID	NT AA Score Probability
ORF Name 33242938t233		- Score Propability
3		Length Length Score Probability
33242938 <u>t2</u> 33		Length Length Score Propability  456 1371 433 1.1e-40
33242938 <u>t2</u> 33		Length Length Score Probability  456 1371 433 1.1e-40  Locus Name Acc#
Protein name  Description		Length Length Score Probability  456 1371 433 1.1e-40  Locus Name Acc#
Protein name  Description	5282	Length         Length         Score         Probability           456         1371         433         1.1e-40           Locus Name         Acc#           sp:Y076_SYNY3         Q55792
Protein name  Description	5282	Length Length Score Probability  456 1371 433 1.1e-40  Locus Name Acc#
Protein name  Description  HYPOTHETICAL 50.0	KD PROTEIN SLR0076  NTID AAID	Length         Length         Score         Probability           456         1371         433         1.1e-40           Locus         Name         Acc#           sp:Y076         SYNY3         Q55792    NT  AA  Score  Probability
Protein name  Description  HYPOTHETICAL 50.0	KD PROTEIN SLR0076  NTID AAID	Length Length Score Probability  456 1371 433 1.1e-40  Locus Name Acc#  sp:Y076 SYNY3 Q55792  NT AA Length Length Score Probability
Protein name  Description  HYPOTHETICAL 50.0  ORF Name  33448342_f1_8	KD PROTEIN SLR0076  NTID AAID	Length         Length         Score         Probability           456         1371         433         1.1e-40           Locus         Name         Acc#           sp:Y076         SYNY3         Q55792           NT         AA         Score         Probability           Length         Length         Probability

ORF Name NTID AAID Length Length Probability
35442313 f1 1 62 5284 292 879 864 2.4e-86
Tagus Nome Acet
Procein name
sp:FUCO_ECOLI P11549
Description
LACTALDEHYDE REDUCTASE, (PROPANEDIOL OXIDOREDUCTASE)
NT AA Score Probability
ORF Name NTID AAID Length Length
3942813_f3_47 [63] [5285] [215] [648] [1026] [1.7e-103]
Protein name Locus Name Acc#
L-fuculose-1-phosphate aldolase gp:AF137263 AF137263
Description
Bacteroides thetaiotaomicron 30S ribosomal protein S16-likeprotein, fucose gene cluster, and RNA polymerase sigma factorSigZ-like protein (sigZ) genes, complete cds.
$rac{ ext{NT}}{ ext{CRF Name}} rac{ ext{AA}}{ ext{NTID}} rac{ ext{AAID}}{ ext{AAID}} rac{ ext{Length}}{ ext{Length}} rac{ ext{Score}}{ ext{tore}} rac{ ext{Probability}}{ ext{Probability}}$
4001515_c1_6064 5286 388 1167 135 3.7e-06
Protein name Locus Name Acc#
transmembrane sensor gp:AF051691 AF051691
Description
Pseudomonas aeruginosa stress factor A (psfA), ECF sigma factor(fiuI), transmembrane sensor (fiuR), and hydroxamate-typeferrisiderophore receptor (fiuA) genes, complete cds.
NT AA Score Probability
ORF Name NTID AAID Length Length
4196001 <u></u>
Protein name Locus Name Acc#
L-fuculose kinase gp:AF137263 AF137263
Description
Bacteroides thetaiotaomicron 30S ribosomal protein S16-likeprotein, fucose gene cluster, and RNA polymerase sigma factorSigZ-like protein (sigZ) genes, complete cds.

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
5197137 £1 6	166	5288		3051	1669	] [1.2e-171
Protein name			<u> </u>		L Nome	7.55#
				_	s Name	Acc#
Initiation factor	1F2-alpha			gp:ECA	72540	AJ002540
Description						
Escherichia coli initiation factor		AU9307) ii	ifB gene	encoding	jtranslat	ional
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
5366453_f1_9	67	5289	416	1251	1106	5.5e-112
Protein name				Locus	s Name	Acc#
nifS-like protein				gp:MLCI	322	Z98741
Description				l L		—— <b>l</b>
Mycobacterium lep	rae cosmid E	322.			-	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 6.2520.33t227		<u>AAID</u> 5290	Length		Score 259	Probability  [8.1e-22
			Length	Length		*
6.2520.33 <u></u> .£2 <u></u> .27			Length	Length	259 Name	8.1e-22
6.2520.33f22.7 Protein name			Length	Length 204 Locus	259 Name	8.1e-22 <u>Acc#</u>
6252033_f2_27  Protein name  L-fucose permease	iotaomicron	30S ribos	Length  67  2  Somal pro	Length Locus gp:AF13	259 S Name 37263 S-likepro	8.1e-22  Acc#  AF137263  Stein, fucose
Protein name  L-fucose permease  Description  Bacteroides theta gene cluster, and complete cds.	iotaomicron RNA polymera	30S ribos ase sigma	Length  67  Comal profactorSi	Locus  Gp:AF1:  Stein S16  .gz-like	259 S Name 37263 S-likepro	8.1e-22  Acc#  AF137263  Stein, fucose
Protein name  L-fucose permease  Description  Bacteroides theta gene cluster, and complete cds.  ORF Name	iotaomicron RNA polymera	30S ribos ase sigma	Length  67  Comal profactorSi  NT  Length	Length Locus gp:AF13 stein S16 gZ-like  AA Length	259 37263 3-likeproprotein Score	AF137263  tein, fucose (sigZ) genes,  Probability
Protein name  L-fucose permease  Description  Bacteroides theta gene cluster, and complete cds.	iotaomicron RNA polymera	30S ribos ase sigma	Length  67  Comal profactorSi  NT  Length	Locus  Gp:AF1:  Stein S16  .gz-like	259 3 Name 37263 5-likeproprotein	Acc#  AF137263  tein, fucose (sigZ) genes,
Protein name  L-fucose permease  Description  Bacteroides theta gene cluster, and complete cds.  ORF Name	iotaomicron RNA polymera	30S ribos ase sigma	Length  67  Comal profactorSi  NT  Length	Length Locus gp:AF13 otein S16 gZ-like  AA Length	259 37263 3-likeproprotein Score	AF137263  tein, fucose (sigZ) genes,  Probability
Protein name  L-fucose permease  Description  Bacteroides theta gene cluster, and complete cds.  ORF Name	iotaomicron RNA polymera	30S ribos ase sigma	Length  67  Comal profactorSi  NT  Length	Length Locus gp:AF13 otein S16 gZ-like  AA Length	259 37263 3-likeproprotein Score 633 3 Name	Acc# AF137263  tein, fucose (sigZ) genes,  Probability  7.3e-62
Protein name  L-fucose permease  Description  Bacteroides theta gene cluster, and complete cds.  ORF Name	iotaomicron RNA polymera	30S ribos ase sigma	Length  67  Comal profactorSi  NT  Length	Length Locus gp:AF1: otein S16 gZ-like AA Length Locus	259 37263 3-likeproprotein Score 633 3 Name	Acc#  AF137263  tein, fucose (sigZ) genes,  Probability  7.3e-62  Acc#

ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
[13913887_c2_82	70	5292		085	564	2.7e-74
Protein name	J		<b></b>	Locu	s Name	Acc#
hypothetical prote	In TM0280			pir:F7	2395	F72395
Description						
ODE Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 15660937c3108		5293		783 T		
Protein name	j [		<u> </u>	l Locu	s Name	Acc#
Description						
NO-HIT						
			NTETS	7.7		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
20734637 <u></u> c280	. 72	5294	263	/92	306	3.9e-45
Protein name				Locu	s Name	Acc#
probable pyruvate enzyme, pflC homolo		yase activ	ating	pir:A6	9431	A69431
Description			*			
				7.7		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
21516.933 <u></u> £1 <u></u> 15	73	5295	76	231	l y	
Protein name				Locu	ıs Name	Acc#
Description						
NO-HIT						
			NT	AA	Caaro	Probability
ORF Name	NTID	AAID	Length	Length	Score	
22461007 <u>f3</u> 37	74	5296	235	708	241	2.5e-20
Protein name				Loci	ıs Name	Acc#
probable competence	e protein	ComF		pir:F	75402	F75402
Description			,			

			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	SCOTE	Hobability
23437627_f1_1	75	5297	88 2	67		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			2777	7) 7)		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
29350677 <u></u> c2 <u></u> 79	76	5298	1368	107	468	1.7e-40
Protein name				Locus	Name	Acc#
bZIP histidine kin	ase			gp:PPU	718245	Y18245
Description						
Pseudomonas putida todI, todH, todS, t	todX, todF	, todC1,	todC2, t	odB, too	IA, todD	,todE, todG,
coar, coun, coas, c	our genes.					
toai, toah, toas, t	odi genes.			7.7		
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
	NTID	<u>AAID</u>	Length		Score	Probability
ORF Name	NTID		Length	Length 210	Score Score	Probability  Acc#
ORF Name 3.05.73.761f14	NTID		Length	Length 210	s Name	
ORF Name 3.05.73.761f14  Protein name	NTID		Length	Length 210	s Name	Acc#
ORF Name  3.05.73.761f14  Protein name  Description	NTID		Length	Length 210  Locus	s Name	Acc#
ORF Name  3.05.73.761f14  Protein name  Description	NTID		Length	Length 210	s Name	Acc#
ORF Name  3.05.7.3.7.6.1	NTID 77 NTID	5299	Length  69  NT Length	Length Locus	s Name	Acc#
ORF Name  30573761_f1_4  Protein name  Description  NO-HIT  ORF Name	NTID  NTID	5299 AAID	Length  69  NT Length	Length Locus  AA Length	s Name	Acc#
ORF Name  3.05.73.761f14  Protein name  Description  NO-HIT  ORF Name  3.622533.7f219	NTID  NTID	5299 AAID	Length  69  NT Length	Length Locus  AA Length	Score	Acc# Probability
ORF Name  3.05.7.3.7.61f14  Protein name  Description  NO-HIT  ORF Name  3.62253.3.7f219	NTID  NTID	5299 AAID	Length  69  NT Length	Length Locus  AA Length	Score	Acc# Probability

ORF Name	NTID AAID	NT AA Score Probability Length Length	
36585962_f3_35	79 5301	418 1257 722 2.7e-71	
Protein name		Locus Name Acc#	
alpha galactosida	se precursor	gp:AF061331 AF061331	
Description			
Saccharopolyspora complete cds.	erythraea alpha ga	lactosidase precursor (melA)gene,	
		NT AA Grand Brochshilita	
ORF Name	NTID AAID	Length Length Score Probability	
3940877_c1_67	80 5302	96 291 163 4.7e-12	
Protein name		Locus Name Acc#	
		sp:YCNE_BACSU P94425	
Description			
HYPOTHETICAL 10.9	KD PROTEIN IN PHRO	C-GDH INTERGENIC REGION	1
			_
		NITT A.A.	]
ORF Name	NTID AAID	NT AA Score Probability	<u>.</u>
ORF Name	NTID AAID	NT AA Score Probability	
	NTID AAID	NT AA Score Probability	
4094627t117 Protein name	NTID AAID	NT AA Score Probability Length Length 879  Locus Name Acc#	
Protein name 115K outer membra	NTID <u>AAID</u>	NT AA Score Probability Length Length 879  Locus Name Acc#	
Protein name  115K outer membra protein	NTID <u>AAID</u>	NT AA Score Probability Length Length 879  Locus Name Acc#  pir:JC6027  JC6027	
Protein name  115K outer membra protein	NTID <u>AAID</u>	NT AA Length Score Probability  1060 3183 879 6.3e-88  Locus Name Acc#  pr:SusC pir:JC6027 JC6027	
Protein name  115K outer membra protein  Description	NTID AAID  81 5303  Ine protein precurso  NTID AAID	NT AA Score Probability  Length Length Score Probability  Locus Name Acc#  Dr:SusC pir:JC6027 JC6027	
Protein name  115K outer membra protein  Description  ORF Name	NTID AAID  81 5303  Ine protein precurso  NTID AAID	NT AA Length Score Probability  Locus Name Acc#  Dr:SusC pir:JC6027 JC6027  NT AA Length Score Probability	
Protein name  115K outer membra protein  Description  ORF Name  4101507_c3_101	NTID AAID  81 5303  The protein precurso  NTID AAID  NTID AAID  82 5304	NT Length         AA Length         Score         Probability           1060         3183         879         6.3e-88           Locus Name         Acc#           pir:JC6027         JC6027           NT Length         AA Length         Score         Probability           182         1149         842         5.2e-84	
Protein name  115K outer membra protein  Description  ORF Name  4101507_c3_101	NTID AAID  81 5303  The protein precurso  NTID AAID  NTID AAID  82 5304	NT Length         AA Length         Score         Probability           1060         3183         879         6.3e-88           Locus Name         Acc#           Dr:SusC         pir:JC6027         JC6027           NT Length         Length         Score         Probability           Length         Length         842         5.2e-84           Locus Name         Acc#	

ORF Name	NTID AAID Length Length Score Probability
4103812 f2_21	83   5305   271   816   360   6.2e-33
Protein name	Locus Name Acc#
	sp:SUHB_ECOLI
Description	
EXTRAGENIC SUPPRI	SSOR PROTEIN SUHB
ORF Name	$rac{ ext{NT}}{ ext{NTID}} = rac{ ext{AA}}{ ext{AAID}} = rac{ ext{Eength}}{ ext{Length}} = rac{ ext{Score}}{ ext{Score}} = rac{ ext{Probability}}{ ext{Probability}}$
4422768_f1_18	84 5306 95 288
Protein name	Locus Name Acc#
Description	
NO-HIT	
ORF Name	$rac{ ext{NT}}{ ext{NTID}}$ $rac{ ext{AA}}{ ext{AAID}}$ $rac{ ext{Length}}{ ext{Length}}$ $rac{ ext{Score}}{ ext{Core}}$ $rac{ ext{Probability}}{ ext{Probability}}$
ORF Name 4534660c162	NTID AAID Length Length Score Propability
	NTID AAID Length Length Score Propability
4534660 <u>c1_</u> 62	NTID AAID Length Length Score Probability 85 5307 443 1332 620 5.5e-78
Protein name  Description	NTID AAID Length Length Score Probability  85   5307   443   1332   620   5.5e-78    Locus Name   Acc#   Sp:XYLE_ECOLI   P09098
Protein name  Description	NTID AAID Length Length Score Probability 85 5307 443 1332 620 5.5e-78  Locus Name Acc#
Protein name  Description	NTID AAID Length Length Score Probability  85   5307   443   1332   620   5.5e-78    Locus Name   Acc#   Sp:XYLE_ECOLI   P09098
4534660c162  Protein name  Description  D-XYLOSE-PROTON	NTID AAID Length Length Score Probability  85   5307   443   1332   620   5.5e-78    Locus Name   Acc#    Sp:XYLE_ECOLI   P09098    YMPORTER (D-XYLOSE TRANSPORTER)  NTID   AAID   Length Length   Score   Probability    NTID   AAID   Length Length   Length   Length   Length    NTID   AAID   Length   Length   Length   Length    NTID   AAID   Length   Length   Length   Length    NTID   AAID   Length   Length   Length    NTID   AAID   Length   Length   Length    NTID   AAID   Length    NTID   AAID
Protein name  Description  D-XYLOSE-PROTON  ORF Name	NTID AAID Length Length Score Probability  85   5307   443   1332   620   5.5e-78    Locus Name   Acc#    Sp:XYLE_ECOLI   P09098    YMPORTER (D-XYLOSE TRANSPORTER)  NTID   AAID   Length Length   Score   Probability    NTID   AAID   Length Length   Length   Length   Length    NTID   AAID   Length   Length   Length   Length    NTID   AAID   Length   Length   Length   Length    NTID   AAID   Length   Length   Length    NTID   AAID   Length   Length   Length    NTID   AAID   Length    NTID   AAID
Protein name  Description  D-XYLOSE-PROTON  ORF Name  4799033t230	NTID AAID Length Length Score Probability    NTID   AAID   Length   Length   Score   Probability

ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score E	robability
10820130_c1_218	87 5309	68 207		
Protein name		Locus	Name	Acc#
Description				
NO-HIT				
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score I	Probability
11.723.43.7c33.54	88 5310	79 240		<b>7</b> !!
Protein name		Locus	Name	Acc#
<u>Description</u>				
NO-HIT				
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score I	Probability
1207152 <u></u> c3 <u></u> 351	89 5311	162 489	217	8.9e-18
Protein name		Locus		Acc#
unknown		gp:AF12	5164	AF125164
Description	lis 638R polysaccha	ride B (PS B2) bio	svnthesis	locus,
	and unknown genes.	1146 5 (15 52) 520	7	
		NT AA	Score	Probability
ORF Name	NTID AAID	Length Length		
12754816 <u></u> £2 <u>.</u> .114	90 5312	209 630		
Protein name		Locus	Name	Acc#
Description		5		
NO-HIT				

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
14296885_c1_248  Protein name	91	5313	200 6	03 Locus	260 Name	2.5e-22 Acc#
conserved hypotheti	.cal protei	n AF0781		pir:E69	9347	E69347
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
146.26.317 <u></u> ±1 <u></u> 52	92	5314	640	.923	736	1.0e-78
Protein name				gp: ECU	Name 19166	<u>Acc#</u> U89166
Description						V (4)
Eikenella corrodens	s lysine de	carboxyla	ase (ECOR	LD) gene	e, comple	tecds.
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 14648562t2116	<u>NTID</u>	<u>AAID</u> 5315	Length		Score 889	Probability  5.5e-89
14648562t2116 Protein name	93	5315	Length 579	Length 740 Locus	889 Name	5.5e-89 <u>Acc#</u>
	93	5315	Length 579	Length	889 Name	5.5e-89
14648562_t2_116 Protein name single-strand DNA-s	93	5315	Length 579	Length 740 Locus	889 Name	5.5e-89 <u>Acc#</u>
Protein name single-strand DNA-shomolog yrvE	93	5315	Length 579	Length 740 Locus	889 Name	5.5e-89 <u>Acc#</u>
Protein name single-strand DNA-shomolog yrvE Description	pecific ex	5315 conuclease	Length  579  NT Length	Length  Locus  pir:H63  AA  Length	889 Name 9980	5.5e-89 Acc# H69980
Protein name single-strand DNA-shomolog yrvE Description  ORF Name  15.735882_c1_208	NTID  94  ein-related	AAID	Length  579  NT  Length  400  1	Length  Locus  pir:H63  AA  Length	Score 696 Name	5.5e-89  Acc#  H69980  Probability  1.5e-68

ORF Name	NTID AAID	NT AA Length Length Score	Probability
16016075_c2_288	95 5317	85 258	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Length Length Score	Probability
16616302f3197	96 5318	439 1320 1138	2.3e-115
Protein name		Locus Name	Acc#
coenzyme F390 synth	netase (ftsA-3) hom	olog pir:D69501	D69501
Description			
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> <u>Score</u> <u>Length</u> <u>Length</u>	Probability
16835312 <u></u> £1 <u></u> 34	97 5319	279 840	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
		3.700 B.70	
ORF Name	NTID AAID	NT AA Length Length Score	Probability
192693 <u></u> c2289	98 5320	211 636 236	1.4e-19
Protein name		Locus Name	Acc#
		sp:YHCG ECOLI	P45423
Description		NANT INTERGENIC REGION (	

			NT	AA	_		
ORF Name	NTID	AAID	Length	Length	Score	Probability	
19535652_c3_316	99	5321	640	.923	940	2.2e-94	]
Protein name				Locus	s Name	Acc#	
putative epimera	se/dehydratas	e WbiI		gp:AF06	54070	AF064070	
Description							
Burkholderia pse putative 1-acyl-s diadenosine tetra biosynthesis gene phosphateN-acetyl	n-glycerol-3- phosphatase(a cluster,comp	phosphate paH), cor lete sequ	eacyltran mplete co uence; pu	sferase ls; type tative u	(plsC), II O-ant	putative tigen	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
2031552 <u></u> c2 <u></u> 298	100	5322	101 3	06	74	0.013	]
Protein name				Locus	Name	Acc#	
				sp:NU3N	_RAT	P05506	
Description							
NADH-UBIQUINONE	OXIDOREDUCTAS	E CHAIN	3,		:		
ORF Name	NTID	E CHAIN 3	NT Length	AA Length	Score	Probability	
	NTID		NT Length		Score	Probability	
ORF Name	NTID	AAID	NT Length	Length	Score .	Probability  Acc#	
ORF Name 20516500 <u></u> c2304	NTID	AAID	NT Length	Length			
ORF Name 2.0516500_c2_304  Protein name	NTID	AAID	NT Length	Length			
ORF Name  2.0516500_c2_304  Protein name  Description	NTID	AAID	NT Length 705 2	Length  118  Locus			*
ORF Name  2.0516500_c2_304  Protein name  Description	NTID	AAID	NT Length	Length		* , ,	
ORF Name  2.0516500c2304  Protein name  Description  NO-HIT	NTID  101  NTID	<u>AAID</u> 5323	NT Length 705 2 NT Length	Length  Locus	s Name	Acc#	
ORF Name  2.05165.0.0c23.04  Protein name  Description  NO-HIT  ORF Name	NTID  101  NTID	<u>AAID</u> 5323 <u>AAID</u>	NT Length  705  2  NT Length	Length  Locus  AA  Length  386	Score	Acc# Probability	]
ORF Name  2.05165.00c23.04  Protein name  Description  NO-HIT  ORF Name  211687c2299	NTID  101  NTID	<u>AAID</u> 5323 <u>AAID</u>	NT Length  705  2  NT Length	Length  Locus  AA  Length  386	Score 309 Name	Acc# Probability 2.0e-27	
ORF Name  2.05165.00c23.04  Protein name  Description  NO-HIT  ORF Name  211687c2299  Protein name	NTID  101  NTID	<u>AAID</u> 5323 <u>AAID</u>	NT Length  705  2  NT Length	Length  Locus  AA  Length  386  Locus	Score 309 Name	Acc#  Probability  2.0e-27  Acc#	]_

		$\frac{NT}{T}$ , $\frac{AA}{T}$ , Sco	re Probability
ORF Name	NTID AAID	Length Length	
2117305_c3_317	103 5325	856 2571 464	2.7e-56
Protein name		Locus Nam	e <u>Acc#</u>
otnA protein		pir:S70958	\$70958
Description	Ē		
ORF Name	NTID AAID	NT AA Sco Length Length	re <u>Probability</u>
21485952 <u></u> c1 <u></u> 244	104 5326	965 2898 821	1.0e-120
Protein name		Locus Nam	e Acc#
		sp:YDIJ_ECO	LI P77748
Description			
HYPOTHETICAL 113.	2 KD PROTEIN IN LP	P-AROD INTERGENIC REGIO	N
ORF Name	NTID AAID	<u>NT AA</u> Sco Length Length	
	DIAA DITM	NT AA SCO	
ORF Name	DIAA DITM	<u>NT</u> <u>AA</u> Sco Length Length	re Probability
ORF Name 21500933t141	DIAA DITM	NT AA Score	re Probability
ORF Name 21500933_t1_41	DIAA DITM	NT AA Score	re Probability
ORF Name  21500933_t1_41  Protein name  Description	DIAA DITM	NT AA Scone Length Scone Locus Nam	re Probability e Acc#
ORF Name  21500933_t1_41  Protein name  Description	DIAA DITM	NT AA Score	e Acc#
ORF Name  21500933t141  Protein name  Description  NO-HIT	NTID AAID 105   5327  NTID AAID	NT AA SCO	re Probability e Acc#
ORF Name  21500933t141  Protein name  Description  NO-HIT  ORF Name	NTID AAID 105   5327  NTID AAID	NT AA Score	re Probability  e Acc#  re Probability
ORF Name  21500933t141  Protein name  Description  NO-HIT  ORF Name  22345000t288	NTID AAID 105   5327  NTID AAID	NT AA Score	re Probability  e Acc#  re Probability

ORF Name	NTID AAII	NT AA D Length Lengt	h Score	Probability
22551562_t1_48	107   5329	143 432	7	
Protein name		Loc	cus Name	Acc#
Description				
NO-HIT				
		NT AA	Canada	Drobobilit
ORF Name	NTID AAII	Length Lengt	h Score	Probability
23546952 <u>t2</u> 106	108 5330	291 876	]	
Protein name		Loc	us Name	Acc#
Description				
NO-HIT		-		
ORF Name	NTID AAII	NT AA Length Lengt	h Score	Probability
2356.718.7 <u></u> £2 <u></u> 102	109 5331	109 330		
Protein name		Loc	us Name	Acc#
Description				
NO-HIT				
		NT AA		
ORF Name	NTID AAII	<del></del>	score	Probability
236.09457f127	110 5332	64 195	]	
Protein name		Loc	us Name	Acc#
Description				
NO-HIT			-	
		NIT 7.7		
ORF Name	NTID AAII	<u>NT</u> <u>AA</u> D <u>Length</u> <u>Lengt</u> l	score	Probability
23.633592 <u>c1</u> 237	111 5333	64 195	]	
Protein name		Loc	us Name	Acc#
Description				
NO-HIT				

ORF Name	NTID A	NT AID Length	$\frac{AA}{Length}$ Score	Probability	
24103388_£1_71	112 5334	531 [1	596 831	7.7e-83	
Protein name			Locus Name	Acc#	
indolepyruvate	oxidoreductase, al	pha subunit	pir:G69114	G69114	
Description					
		NT	7.7		
ORF Name	NTID A	AID Length	AA Score	Probability	
243.03127 <u></u> f1 <u></u> 33		142 4	29		
Protein name			Locus Name	Acc#	
Description					
NO-HIT					1
		· · · · · · · · · · · · · · · · · · ·			J
		NŒ	7.7		
ORF Name	NTID AF	<u>NT</u> AID Length	AA Length Score	Probability	
ORF Name 24318802_f3_187		AID Length	— Score	Probability	
		AID Length	Length Score		
24318802 <u></u> £3 <u></u> 187		AID Length	Length Score 233 490	1.0e-46	
24318802 <u></u> £3 <u></u> 187		AID Length	Locus Name	1.0e-46 <u>Acc#</u>	
24318802f3187 Protein name		AID Length	Locus Name	1.0e-46 <u>Acc#</u>	
Protein name  Description		AID Length 410 [1	Length Score  Length 490  Locus Name  Sp:XYLR_ANATH	1.0e-46 Acc# Q44406	
Protein name  Description		AID Length	Locus Name	1.0e-46 <u>Acc#</u>	
Protein name  Description  XYLOSE REPRESSO	R NTID AA	AID Length  AID T  AID Length  NT  Length	Length Score  Length Score  233 490  Locus Name  Sp:XYLR_ANATH	1.0e-46 Acc# Q44406	
Protein name  Description  XYLOSE REPRESSO  ORF Name	R NTID AA	AID Length  AID T  AID Length  NT  Length	Length Score  Length Score  233 490  Locus Name  Sp:XYLR_ANATH  AA  Length Score	1.0e-46 Acc# Q44406	
Protein name  Description  XYLOSE REPRESSO  ORF Name  24402177_f2_103	R NTID AA	AID Length  AID T  AID Length  NT  Length	Length Score  Length Score  233 490  Locus Name  Sp:XYLR_ANATH  AA Length Score  85	1.0e-46 Acc# Q44406  Probability	

			NT	AA	Score	Probability	
ORF Name	NTID	AAID L	ength	Length	<u>30016</u>	FIODADITIC	<u> </u>
24415908_f1_72	116 53	38 1	95 58	88	310	1.2e-27	
Protein name				Locus	Name	Acc#	
indolepyruvate fer subunit beta (iorB)		reductase	€,	pir:E69	503	E69503	
Description							
ORF Name	NTID	AAID L	NT ength	<u>AA</u> Length	Score	Probability	7
24415930 <u></u> c1 <u></u> 239	117 53	39 3:	22 96	9	937	4.5e-94	
Protein name				Locus	Name	Acc#	
WbpB				gp:PAU5	0396	U50396	
Description			_				
Pseudomonas aerugi WbpB (wbpB), WbpC ( Wzx (wzx), HisH (hi WbpJ (wbpJ), WbpK ( complete cds, and U	wbpC), WbpD ( sH), HisF (hi wbpK), WbpL(w	wbpD), Wk sF), WbpG bpL), Wbr	opE (wbp ,(wbpG), oM (wbpM	E),Wzy WbpH (	(Rfc) (w wbpH), W	zy (rfc)), bpI (wbpI),	
ORF Name	NTID	AAID L	<u>NT</u> ength	<u>AA</u> Length	Score	Probability	7
2442633.7 <u></u> .c1 <u>.</u> .216	118 53	40 8:	9 27	0			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT					-		
ORF Name	NTID	AAID L	<u>NT</u> ength	<u>AA</u> Length	Score	Probability	-
246408.75 <u></u> c1 <u></u> 245	119 53	41 3	76 11	31	1914	1.3e-197	*
Protein name			v	Locus	Name	Acc#	
putative aminotran	sferase			gp:AF12	5164	AF12516	54
Description							
Bacteroides fragil complete sequence;			B (PS	B2) bio	synthesi	slocus,	-

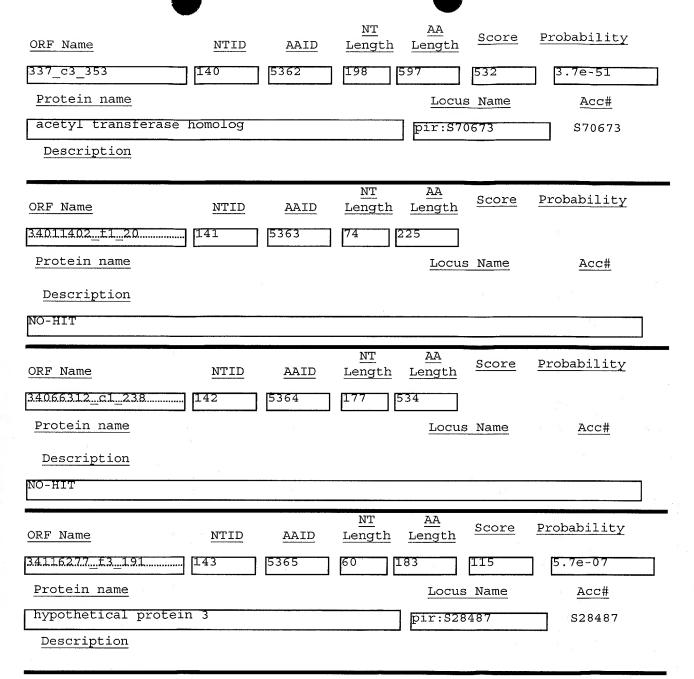
· ·	<b>-</b> '		
ORF Name	NTID AAID	NT AA Length Length Score	Probability
24651562_t1_68	120 5342	522 1569 318	3.1e-25
Protein name		Locus Name	Acc#
surface antigen Bsp	)A	pir:T31094	T31094
Description			
ORF Name	NTID AAID	NT AA Length Length Score	Probability
24806502 <u></u> c1249	121 5343	325 978 581	2.4e-56
Protein name		Locus Name	Acc#
		sp:FMT_BACSU	
Description			
METHIONYL-TRNA FORM	MYLTRANSFERASE,		1
		2.77	
ORF Name	NTID AAID	NT AA Length Length Score	Probability
25421942 <u></u> c1207	122 5344	181 546 152	6.9e-11
Protein name		Locus Name	Acc#
unknown		gp:AF048749	AF048749
Description	*		
Bacteroides fragil: sequence.	is capsular polysac	charide biosynthesis ope	ron, complete
		NT AA	
ORF Name	NTID AAID	Length Length Score	Probability
25429812 <u>f3</u> 171	5345	341 1026 111	0.0022
Protein name		Locus Name	Acc#
		sp:Y973_METJA	Q58383
Description			
HYPOTHETICAL PROTE	IN MJ0973		

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
25583577_c3_356	124	5346		182	172	3.4e-10
Protein name	_  <del>  </del>		L.	Locus	Name	Acc#
Cap5I				gp:SAU	31973	U81973
Description				L		
Staphylococcus aur complete cds.	eus capsule	gene cli	ıster Cap	5A throu	igh Cap5P	genes,
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
25667592_c3_363	125	5347	606 1	821	537	4.0e-55
Protein name				Locus	Name	Acc#
chloride channel,	probable, h	omolog		pir:F69	9426	F69426
Description						
			NT	AΑ		
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 25676387c3352	<del></del>	<u>AAID</u> 5348	Length	*********	Score	Probability 2.8e-07
	<del></del>		Length	Length 52		
25676387 <u></u> 3. <u></u> 3.52	<del></del>		Length	Length 52	118 Name	2.8e-07
256.7638.7c3352 Protein name	<del></del>		Length	Length 52 Locus	118 Name	2.8e-07 <u>Acc#</u>
256.7638.7c3352 Protein name tachylectin-3	126	5348	Length  83  2	Length  52  Locus  gp:AB01	118 Name 7484	2.8e-07 <u>Acc#</u>
256.7638.7c3352  Protein name tachylectin-3 Description	126	5348	Length  83  2	Length  52  Locus  gp:AB01	118 Name 7484	2.8e-07 <u>Acc#</u>
256.763.8.7c33.52	atus mRNA fo	5348 or tachyl	Length  83  2  ectin-3,  NT  Length	Length  52  Locus  gp:AB01  complet	118 S Name .7484 e cds.	2.8e-07 <u>Acc#</u> AB017484
256.7638.7c3352	atus mRNA fo	or tachyl	Length  83  2  Lectin-3,  NT  Length	Length  52  Locus  gp:AB01  complete  AA  Length	118 S Name 7484 Se cds. Score	2.8e-07  Acc#  AB017484  Probability
256.7638.7c3352  Protein name tachylectin-3 Description Tachypleus trident  ORF Name 259.76510c2263	atus mRNA fo	or tachyl	Length  83  2  Lectin-3,  NT  Length	Length  52  Locus  gp:AB01  complete  AA  Length	118 S Name -7484 Se cds. Score 271 S Name	2.8e-07  Acc#  AB017484  Probability  1.7e-23
256.7638.7c3352  Protein name tachylectin-3 Description Tachypleus trident  ORF Name 259.76510c2263	atus mRNA fo	or tachyl	Length  83  2  Lectin-3,  NT  Length	Length  Locus  Gp:AB01  complete  AA  Length  36	118 S Name -7484 Se cds. Score 271 S Name	2.8e-07  Acc#  AB017484  Probability  1.7e-23  Acc#
Protein name  tachylectin-3  Description  Tachypleus trident  ORF Name  2597.6510c2263	atus mRNA fo	or tachyl AAID	ectin-3,  NT Length  311  9	Length  52  Locus  gp:AB01  complet  AA  Length  36  Locus  gp:ECNE	118 S Name 7484 Se cds. Score 271 S Name	2.8e-07  Acc#  AB017484  Probability  1.7e-23  Acc#  X03345

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
26604635_c3_355	128	5350	370	1113	678	1.3e-66	
Protein name				Locus	s Name	Acc#	
unknown				gp:AF1	44879	AF14487	9
Description				-			
Leptospira interro	gans rfb lo	cus, com	plete se	quence.			1
			NTITI	7. 7.			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
29305313_c1_242	129	5351	490	1473	122	0.00028	
Protein name				Locus	s Name	Acc#	
putative polysacch	aride polym	erase		gp:SPU(	9239	U09239	
Description							
Streptococcus pneu operon, (cps19fABCD		8 -	_		_		
cds.		,, genes,	complete	e cds, and	alia ge	ene, partial	
cds.		,, genes,	NT	AA			
ORF Name	NTID	AAID	1		Score	Probability	-
	NTID		<u>NT</u> Length	AA			-
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length			-
ORF Name 29317660_c1_204	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name  29317660_c1_204  Protein name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name  29317660_c1_204  Protein name  Description	NTID	AAID	NT Length	AA Length 378 Locus	Score	Probability	
ORF Name  29317660_c1_204  Protein name  Description	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name  29317660c1204  Protein name  Description  NO-HIT	NTID  NTID	<u>AAID</u> 5352	NT Length  125	AA Length  B78  Locus	Score Name	Probability  Acc#	
ORF Name  29317660_c1_204  Protein name  Description  NO-HIT  ORF Name	NTID  NTID	<u>AAID</u> 5352 <u>AAID</u>	NT Length  125	AA Length Locus  AA Locus  AA Length	Score Name	Probability  Acc#	
ORF Name  29317660c1204  Protein name  Description  NO-HIT  ORF Name  29509630c1217	NTID  NTID	<u>AAID</u> 5352 <u>AAID</u>	NT Length  125	AA Length Locus  AA Locus  AA Length	Score  Score	Probability  Acc#  Probability	

30355312_c1_246	
DNA-binding protein HB pir:C75600 C75600  Description	
31381c1209133 5355 416 1251 289 6.5e-24	
Protein name Locus Name Acc#	
sp:YYBO_BACSU P37489	
Description	
HYPOTHETICAL 48.2 KD PROTEIN IN COTF-TETB INTERGENIC REGION	
ORF Name NTID AAID Length Length Probability	
31414025c2290	
Protein name Locus Name Acc#	
sp:3MG1_ECOLI P05100	
Description	
$\frac{\text{NT}}{\text{NT}}$ , $\frac{\text{AA}}{\text{AA}}$ , Score Probability	
ORF Name NTID AAID Length Length	
31882036c2271	
Protein name Locus Name Acc#	
Description	
NO-HIT	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
32064137_c2_294	136	5358	209	530	120	2.3e-07
Protein name				Locus	Name	Acc#
unknown				gp:AF04	18749	AF048749
Description						
Bacteroides fragil sequence.	lis capsular	polysaco	charide k	oiosynthe	esis ope	ron, complete
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
32323912_f1_17	137	5359	411 1	236		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT					21	
110-1111						
110-1111			NT	7.7		
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
		<u>AAID</u>	<u>Length</u>		Score	Probability 1.1e-18
ORF Name			<u>Length</u>	Length	i	
ORF Name 32429512f2113	138	5360	<u>Length</u>	Length	163 Name	1.1e-18
ORF Name  32429512f2113  Protein name	138	5360	<u>Length</u>	Length 2316 Locus	163 Name	1.1e-18 Acc#
ORF Name  32429512_f2_113  Protein name  DNA repair protein	138	5360	<u>Length</u>	Length 2316 Locus	163 Name 9294	1.1e-18 Acc# F69294
ORF Name  32429512_f2_113  Protein name  DNA repair protein	138	5360	Length	Length  316  Locus  pir:F69	163 Name	1.1e-18 Acc#
ORF Name  32429512_f2_113  Protein name  DNA repair protein  Description	nTID	10g	Length  771  2  NT Length	Length  2316  Locus  pir:F69	163 Name 9294	1.1e-18 Acc# F69294
ORF Name  32429512_f2_113  Protein name  DNA repair protein  Description  ORF Name	nTID	log  AAID	Length  771  2  NT Length	Length  Locus  pir:F69  AA  Length	163 Name 9294	1.1e-18 Acc# F69294
ORF Name  32429512f2113  Protein name  DNA repair protein  Description  ORF Name  32617177f295	nTID	log  AAID	Length  771  2  NT Length	Length  Locus  pir:F69  AA  Length	163 Name 9294 Score	1.1e-18  Acc#  F69294  Probability



ODE Nama	NUTTO AATO	NT Longth	AA Score Probabil	ity
ORF Name	NTID AAID	Length		
34167567_c2_297	144 5366	446	1341 1226 1.1e-1:	24
Protein name			Locus Name Acc	<u>#</u>
ORF1P			gp:AB025970 AB02	5970
Description				
<del>_</del>		1P, ORF21	P, ORF3P, ORF4P, ORF5P, ORF	6P,
ORF7P, ORF8P, ORF91	P, ORF10P, ORF11P.			00
		NT	AA Score Probabil:	
ORF Name	NTID AAID	<u>Length</u>	Length Score Probabil:	ILLY
36361063_c2_300	145 5367	370	610 2.0e-5	9
Protein name			Locus Name Acc	<u>#</u>
WbpH			gp:PAU50396 U503	96
Description	· ·			
WbpB (wbpB), WbpC Wzx (wzx), HisH (hi WbpJ (wbpJ), WbpK	(wbpC), WbpD (wbpD), isH), HisF (hisF), W	WbpE (wk (Dqdw)BqdW 1dw) MqdW	gene, partial cds,WbpA (wb ppE),Wzy (Rfc) (wzy (rfc)) , WbpH (wbpH), WbpI (wbpI pM) and WbpN (wbpN) genes,	,
		NT	AA Score Probabil:	itv
ORF Name	NTID AAID	<u>Length</u>	Length	
3911015 <u></u> c3 <u></u> 358	146 5368	180	543 153 5.4e-1	
Protein name			Locus Name Acc	<u>#</u>
serine O-acetyltra	nsterase,	-	pir:E53402 E534	02
Description				
ORF Name	NTID AAID	<u>NT</u> Length	AA Score Probabil:	ity
3948587 <u></u> c2 <u></u> 296	147 5369	399	.200 1198 9.9e-12	22
Protein name			Locus Name Acc	<u>#</u>
			gp:D64132 D641	32
Description				
Porphyromonas ging	jivalis PorR and Por	S genes,	complete cds.	*

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4022178_c2_262	148	5370	373	122	86	7.1e-07	
Protein name				Locus	s Name	Acc#	
				sp:YCC	C_ECOLI	_	
Description				<u> </u>	-		
HYPOTHETICAL 81.2	KD PROTEIN	IN APPA-C	CSPH INTE	ERGENIC F	REGION		
			NTO	7. 7.			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4033125_c1_250	149	5371	226	81	546	1.2e-52	
Protein name				Locus	Name	Acc#	
ribulose-5-phospha	te 3-epimer	ase homol	og yloR	pir:B69	9879	B69879	
Description							
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
ORF Name 40.656.7.7£213.3		<u>AAID</u> 5372	Length		Score	Probability 9.0e-25	
		<del></del>	Length	Length			
40.656.7.7±213.3	150	5372	Length	Length	283 Name	9.0e-25	
4065677f2133 Protein name	150	5372	Length	Length 047 Locus	283 Name	9.0e-25 <u>Acc#</u>	
Protein name  conserved hypothet	150	5372	Length  348	Length 047  Locus pir:D70	283 Name	9.0e-25 <u>Acc#</u> D70188	
Protein name  conserved hypothet	150	5372	Length	Length 047 Locus	283 Name	9.0e-25 <u>Acc#</u>	
Protein name  Conserved hypothet  Description	ical protein	5372 n BB0709	Length  348 [1]  NT  Length	Length  047  Locus  pir:D70	283 S Name	9.0e-25 <u>Acc#</u> D70188	
Protein name  Conserved hypothet  Description  ORF Name	ical protein	5372 n BB0709	Length  348 [1]  NT  Length	Length  047  Locus  pir:D70  AA  Length	283 Name 0188 Score	9.0e-25 <u>Acc#</u> D70188  Probability	
Protein name  Conserved hypothet  Description  ORF Name  40.968.77t169	ical protein	5372 n BB0709	Length  348 [1]  NT  Length	Length  047  Locus  pir:D70  AA  Length	283 3 Name 0188 Score 348 3 Name	9.0e-25  Acc#  D70188  Probability  1.2e-31	
Protein name  Conserved hypothet  Description  ORF Name  40.968.77t169	ical protein	5372 n BB0709	Length  348 [1]  NT  Length	Length  047  Locus  pir:D70  AA  Length  91  Locus	283 3 Name 0188 Score 348 3 Name	9.0e-25 <u>Acc#</u> D70188  Probability  1.2e-31 <u>Acc#</u>	

			NT AA	, Score	Probability
ORF Name			ength Leng		
4103375_c1_243	152 53	74 41	1233	843	4.1e-84
Protein name			Lo	cus Name	Acc#
putative transfe	erase		gp:E	BR007747	AJ007747
Description					
Bordetella bron	chiseptica cosmi	d BbLPS1.		-	*
			NT AA		
ORF Name	NTID	AAID Le	ength Length	ch Score	Probability
4457512_f3_184	153 53	75 33	33 1002	221	2.6e-16
Protein name			Lo	cus Name	Acc#
conserved hypoth	netical protein I	МТН83	pir:	F69210	F69210
Description	· ·				
ORF Name	NTID	AAID Le	<u>NT AA</u> ength Lengt	ch Score	Probability
ORF Name 4459660c2252	* "		ength Lengt	ch Score	Probability 9.0e-06
	* "		ength Lengt	<u> </u>	
4459660 <u></u> c2 <u></u> 252		76   26	ength Lengt	109	9.0e-06
Protein name  probable NADH-planbunit		76   26	ength Lengt	109 cus Name	9.0e-06 <u>Acc#</u>
Protein name probable NADH-pl		76   26	ength Lengt	109 cus Name	9.0e-06 <u>Acc#</u>
Protein name  probable NADH-planbunit		76 26	se pir:	109  Cus Name  C71018	9.0e-06 Acc# C71018
Protein name  probable NADH-planbunit	astoquinone oxid	76 26	ength Lengt  9 810  Lo se pir:	109  Cus Name  C71018	9.0e-06 <u>Acc#</u>
Protein name  probable NADH-pl subunit  Description	astoquinone oxio	doreducta	ength Length  Se pir:  NT AA ength Length	109  Cus Name  C71018	9.0e-06 Acc# C71018
Protein name  Probable NADH-pl subunit  Description  ORF Name	astoquinone oxio	doreducta	ength Length  Se pir:  NT AA ength Length  2175	109  Cus Name  C71018	9.0e-06 Acc# C71018  Probability
Protein name  Probable NADH-plandsubunit  Description  ORF Name  4.728385_t3_173		doreducta	se pir:  NT AA Length Length Length Length Length Length Lo	109  Cus Name  C71018  Score  208	9.0e-06 Acc# C71018  Probability 6.3e-13
Protein name  probable NADH-plandsubunit  Description  ORF Name  4.728385_t3_173  Protein name		doreducta	se pir:  NT AA Length Length Length Length Length Length Lo	109  Cus Name  C71018  Score  208  Cus Name	9.0e-06 Acc# C71018  Probability  6.3e-13 Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4786250_£2_83	156	5378	156	471	135	4.3e-09	
Protein name				Locus	s Name	Acc#	
hypothetical protei	n MTH658			pir:E6	9187	E69187	
Description							
			NT	AA	2	December 1, 17, 71	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
4789066f2104	157	5379	71	216			
Protein name				Locus	s Name	Acc#	
Description							
NO-HIT				7447 4 74 74 74 74 74 74 74 74 74 74 74			
			2700	7.7			
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
5359842 <u></u> c3 <u></u> 319	158	5380	174	525			
Protein name				Locus	s Name	Acc#	
Description							
NO-HIT							1
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
58.7.7.042 <u></u> c1251	159	5381	325	75	131	3.0e-05	
Protein name				Locus	Name	Acc#	
				sp:CME3	B_BACSU	P39695	
Description					* 1	-	

ORF Name NTID AAID	NT AA Score Probability
5894001_±1_46	83 252 65 0.020
Protein name	Locus Name Acc#
	Sp:UDG STRPY Q07172
Description	
(UDP-GLCDH) (UDPGDH)	
ORF Name NTID AAID	NT AA Length Length Score Probability
6056562_c3_359 161 5383	204 615 406 8.3e-38
Protein name	Locus Name Acc#
putative transferase	gp:BBR007747 AJ007747
Description	
Bordetella bronchiseptica cosmid BbLM	PS1.
	NT AA
ORF Name NTID AAID	NT AA Score Probability
ORF Name         NTID         AAID           6288313f2105         162         5384	Score Probability
1.	Length Length Score Probability
6288313 <u>f2</u> 105 <u>5384</u>	Length Length Score Probability  308 927 407 6.5e-38
6288313_f2_105162	Length Length Score Probability  308 927 407 6.5e-38  Locus Name Acc#
6288313_f2_105162 5384  Protein name transposase  Description	Length Length Score Probability  308 927 407 6.5e-38  Locus Name Acc#  gp:AF038866 AF038866
Protein name  transposase  Description  Bacteroides fragilis transposon Tn552	Length Length Score Probability  308 927 407 6.5e-38  Locus Name Acc#  gp:AF038866 AF038866  Co transposase (bipH) andmobilization is.
Protein name  transposase  Description  Bacteroides fragilis transposon Tn552	Length Length Score Probability  308 927 407 6.5e-38  Locus Name Acc#  gp:AF038866 AF038866
Protein name  transposase  Description  Bacteroides fragilis transposon Tn552 protein BmpH (bmpH) genes, complete co	Length Length Score Probability    308   927   407   6.5e-38     Locus Name   Acc#     Gp:AF038866   AF038866     20 transposase (bipH) andmobilization     ds.   NT   AA   Score Probability
Protein name  transposase  Description  Bacteroides fragilis transposon Tn552 protein BmpH (bmpH) genes, complete co	Length Length Score Probability    308   927   407   6.5e-38     Locus Name   Acc#     gp:AF038866   AF038866     20 transposase (bipH) andmobilization     ds.   NT
Protein name  transposase  Description  Bacteroides fragilis transposon Tn552 protein BmpH (bmpH) genes, complete co	Length Length Score Probability    308   927   407   6.5e-38     Locus Name   Acc#     gp:AF038866   AF038866     20 transposase (bipH) andmobilization     ds.   NT

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
962777_t1_73  Protein name	164	5386	190	Locus	457 s Name	3.3e-43 Acc#	
Description				sp:XPT	BACSU	P42085	
XANTHINE PHOSPHORIE	BOSYLTRANS	FERASE,					
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
9765913_c1_232	165	5387	67	204	75	0.013	
Protein name				Locus	s Name	Acc#	ì
				sp:HBB_	PANPO	P04244	
Description						9	
	T 3.7						
HEMOGLOBIN BETA CHA	TIN				7		
HEMOGLOBIN BETA CHA	ZIN		NT	AA	Saara	Dvobabilita	]
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
		<u>AAID</u>	Length		Score	Probability	
ORF Name	NTID		Length	Length	Score Name	Probability  Acc#	<u> </u>
ORF Name 10671885_c2_125	NTID		Length	Length		. 6	<u></u>
ORF Name  10671885_c2_125  Protein name	NTID		Length	Length		. 6	
ORF Name  10671885_c2_125  Protein name  Description	NTID		Length  123  3	Length  372  Locus	s Name	Acc#	
ORF Name  10671885_c2_125  Protein name  Description	NTID		Length	Length		. 6	
ORF Name  106.71885c2125  Protein name  Description  NO-HIT	NTID	5388	Length  123  NT Length	Length  372  Locus	s Name	Acc#	
ORF Name  106.718.85c2125  Protein name  Description  NO-HIT  ORF Name	NTID  166  NTID	5388 AAID	Length  123  NT Length	Length  Locus  AA  Length  .008	Score	Acc# Probability	
ORF Name  106.718.85c2125  Protein name  Description  NO-HIT  ORF Name  10.7.35.927c2130	NTID  166  NTID  167	<u>AAID</u>	Length  123  NT Length  335  1	Length  Locus  AA  Length  .008	Score	Acc# Probability  3.2e-07	

pul:
(COL)
1122
4 645
4 -46-9
1100
i un
10
1177
in nan
ar ini
11211
app Triple
p m

ORF Name	NTID A	AID L	NT ength	<u>AA</u> Length	Score	Probability	
10757837_c1_119	168 539	0 4	15 1	248	404	1.4e-37	
Protein name				Locus	: Name	Acc#	
11000211				sp:YRK0	BACSU	P54442	
Description						(4)	
HYPOTHETICAL 46.4	KD PROTEIN IN	BLTR-SP	OIIIC I	NTERGEN	C REGION	4	
			NT	AA		Dechability	
ORF Name	NTID 2	AAID I	Length	Length	Score	Probability	
11953533_f1_20	169 539	91 7	75	228	69	0.0020	
Protein name				Locu	s Name	Acc#	
				sp:HXD	3_BRARE	042370	
Description						61	_
HOMEOBOX PROTEIN	HOX-D3 (FRAGME	NT)	.,				]
			אידי	ΔΔ	×	- 1 1 1 3 2 5	
ORF Name	NTID		<u>NT</u> Length	<u>AA</u> Length	Score	Probability	· · · · ·
ORF Name		AAID	Length		Score	Probability	
•		AAID :	Length	Length	Score	Probability  Acc#	
12541502f251 Protein name		AAID :	Length	Length			
Protein name  Description		AAID :	Length	Length			]
12541502f251 Protein name		AAID :	Length	Length 237 Locu		Acc#	
Protein name  Description  NO-HIT		AAID :	Length	Length  237  Locu	s Name		]
Protein name  Description  NO-HIT  ORF Name		AAID	Length 78	Length  237  Locu	s Name	Acc#	
Protein name  Description  NO-HIT  ORF Name  12693752t264		AAID	NT Length	Length  Locu  AA  Length	s Name	Acc# Probability	
Protein name  Description  NO-HIT  ORF Name  126.937.52t264	NTID 171 53	AAID	NT Length	Length  Locu  AA  Length  627  Locu	Score	Acc#  Probability  1.4e-42  Acc#	
Protein name  Description  NO-HIT  ORF Name  126.93.752t264  Protein name  putative GTP-bir  Description	NTID 53	AAID  AAID  393	NT Length 208	Length  Locu  AA  Length  627  Locu  gp:AT.	Score  451  us Name  AC004786	Acc#  Probability  1.4e-42  Acc#  AC004786	
Protein name  Description  NO-HIT  ORF Name  126.937.52t264	NTID 53	AAID  AAID  393	NT Length 208	Length  Locu  AA  Length  627  Locu  gp:AT.	Score  451  us Name  AC004786	Acc#  Probability  1.4e-42  Acc#  AC004786	]

ORF Name	NTID AAID	NT AA Length Length Score	Probability
14642187_f2_61	172   5394	175  528   511	6.2e-49
Protein name		Locus Name	Acc#
		sp:Y318_HAEIN	P43984
Description			
HYPOTHETICAL PROT	TEIN HI0318		
ORF Name	NTID AAID	NT AA Length Length Score	Probability
15125662_c3_164	173 5395	495 1488 367	1.6e-33
Protein name		Locus Name	Acc#
		gp:D90837	<b>-</b>
Description			<del></del>
E.coli genomic DN	A, Kohara clone #3	47(44.2-44.5 min.).	
-			1
ORF Name	NTID AAID	NT AA Score	Probability
ORF Name 16.907.762_c3_168	NTID AAID	NT AA Score	Probability  6.9e-98
	NTID AAID	NT AA Score	*
16.907.762 <u></u> c3 <u></u> 16.8	NTID AAID	<u>NT AA</u> Length Length Score  360 1083 973	6.9e-98
16.907.762 <u></u> c3 <u></u> 16.8	NTID AAID	NT AA Score Length Length 973  Locus Name	6.9e-98 <u>Acc#</u>
Protein name  Description	NTID AAID	NT AA Length Length Score  360 1083 973  Locus Name  sp:YODE_PSEAE	6.9e-98 <u>Acc#</u>
Protein name  Description	NTID AAID	NT AA Length Length Score  360 1083 973  Locus Name  sp:YODE_PSEAE  E 3'REGION (ORF2)	6.9e-98 <u>Acc#</u>
Protein name  Description	NTID AAID	NT AA Score  Length Length 973  Locus Name  Sp:YODE_PSEAE  AA Score	6.9e-98 <u>Acc#</u>
Protein name  Description  HYPOTHETICAL 40.7	NTID AAID  174 5396  KD PROTEIN IN OPD  NTID AAID	NT AA Score  Length Length 973  Locus Name  Sp:YODE_PSEAE  AA Score	6.9e-98 Acc# Q01609
Protein name  Description  HYPOTHETICAL 40.7	NTID AAID  174 5396  KD PROTEIN IN OPD  NTID AAID	NT AA Score  Length Length 973  Locus Name  sp:YODE_PSEAE  AT AA Score  Length Length Score	Acc# Q01609  Probability
Protein name  Description  HYPOTHETICAL 40.7  ORF Name  20579675_c1_104	NTID AAID  174 5396  KD PROTEIN IN OPD  NTID AAID	NT AA Score  Length Length 973  Locus Name  sp:YODE_PSEAE  NT AA Score  Length Length Score  213 642 496	6.9e-98  Acc#  Q01609  Probability  2.4e-47

# EU
705 11 118
M
350
die stant
FU
in the
類
HOH
अंग महर
11111
49 699
C
5 1

	NT AA Score Probability NTID AAID Length Length	
ORF Name	NTID AAID Length Length  176   5398   388   1167   882   3.0e-88	
20587753_c3_165	Locus Name Acc#	
<u>Protein name</u>	sp:PATB_BACSU Q08432	
Description		
PUTATIVE AMINOTRA	NSFERASE B,	
	NT AA Score Probability	
ORF Name	NTID AAID Length Length	
20601437_£2_65	177 5399 204 615	
Protein name	Locus Name Acc#	
Description		
NO-HIT		<u> </u>
	NT AA Score Probability NTID AAID Length Length	
ORF Name	77.20	
21648312_f2_62	Locus Name Acc#	
Protein name		
Description		1
NO-HIT	7.7	
ORF Name	NT AA Score Probability NTID AAID Length Length	-
22853452 <u></u> c3 <u></u> 156	179   5401   179   540   158   1.6e-11	
Protein name	Locus Name Acc#	
	sp:YP20_BACLI P05332	
Description		
		- 20

ORF Name	NTID AAID	NT AA Score	<u>Probability</u>
22867327_c3_155  Protein name	180 5402	149 450 Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Score	<u>Probability</u>
23492786 <u></u> ±3 <u></u> 6.7	181 5403	70 213	
Protein name		Locus Name	Acc#
Description			
NO-HIT			1 4
		NT AA GGGN	
ORF Name	NTID AAID	Length Length Scor	e <u>Probability</u>
23945302_c2_140	182 5404	230 693 163	4.7e-12
Protein name		Locus Name	Acc#
7		sp:RIBD_METJ	A Q58085
Description			
PUTATIVE RIBOFLAVI	N BIOSYNTHESIS ENZ	YME	
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Scor Length Length	e Probability
24334637c1122	. 183 5405	406 1221 223	3.6e-16
Protein name	<u> </u>	Locus Name	Acc#
cation efflux syste	em (czcB-like)	pir:E70342	E70342
Description	. 1		» 8 (O). 8
ORF Name	NTID AAID	NT <u>AA</u> Scor Length Length	re Probability
24416552c1118	184   5406	123 372 224	1.6e-18
Protein name	J	Locus Name	Acc#
oxidoreductase, al	do/keto reductase	family pir:H72307	H72307
Description			·

			NT	AA		D 1 1 1 2 1 1
ORF Name	$\overline{ ext{NTID}}$	AAID	Length	Length	Score	Probability
26594087_c1_109	185	5407	379	140	896	9.9e-90
Protein name				Locus	3 Name	Acc#
oxidoreductase, a	ldo/keto red	ductase fa	amily	pir:H7	2307	H72307
Description						
		·	NT	ĀA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
3237.755 <u></u> c2132	186	5408	225	578	376	1.3e-34
Protein name				Locus	Name	Acc#
plant-metabolite	dehydrogenas	se homolog	yvgN	pir:C70	0040	C70040
Description						· ·
			2.777	7.7		
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
33392187c1120	187	5409	342 1	.029	608	3.3e-59
Protein name				Locus	Name	Acc#
oxidoreductase, a	ldo/keto red	luctase fa	mily	pir:H72	2307	H72307
Description						
	ė					, ,
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
ORF Name 34198387_c2_134		<u>AAID</u> [5410	Length		Score	Probability
*		*	Length	Length	Score Name	Probability  Acc#
34198387 <u></u> c2 <u>134</u>		*	Length	Length		
34198387c2_134 Protein name		*	Length	Length		
234198387_c2_134 Protein name Description		*	Length	Length 79 Locus		
234198387_c2_134 Protein name Description		*	Length	Length		
Protein name  Description  NO-HIT	188	5410	Length  292  8  NT  Length	Length 79 Locus	s Name	Acc#
34198387_c2_134 Protein name Description NO-HIT ORF Name	188	5410 AAID	Length 292 8	Length  79  Locus  AA  Length	Score	Acc# Probability
34198387_c2_134  Protein name  Description  NO-HIT  ORF Name  35313816_c1_111	NTID	AAID 5411	Length 292 8  NT Length 287 8	Length  79  Locus  AA  Length	Score 761 Name	Acc# Probability  2.0e-75

		NT AA Score Probability
ORF Name	NTID AAID	Length Length [1.1e-56]
3948575_c2_123	190 5412	
Protein name		Locus Name Acc#
		sp:YF08_METJA Q58903
Description		
HYPOTHETICAL ABC	TRANSPORTER ATP-BIN	DING PROTEIN MJ1508
		NT AA Score Probability
ORF Name	NTID AAID	Length Length Score Probability
4064178_f1_25	191 5413	455 1368 197 8.2e-13
Protein name		Locus Name Acc#
aspartate aminotra	ansferase	gp:AF035157 AF035157
Description		
Lactococcus lacti	s aspartate aminoti	ansferase (aspC) gene, completecds.
		NT AA - Darkilita
ORF Name	NTID AAID	NT AA Score Probability Length Length
ORF Name 4335425±398		Score Probability
		Length Length
4335425f398 Protein name	192 5414	Length         Length         Score         Probability           498         1497         370         5.4e-34
4335425t398 Protein name hypothetical prot	192 5414	Length         Length         Score         Probability           498         1497         370         5.4e-34           Locus Name         Acc#
4335425f398 Protein name	192 5414	Length         Length         Score         Probability           498         1497         370         5.4e-34           Locus         Name         Acc#           pir:S75887         S75887
Protein name hypothetical prot Description	in [192] [5414]	Length         Length         Score         Probability           498         1497         370         5.4e-34           Locus         Name         Acc#           pir:S75887         S75887
Protein name  hypothetical prot  Description  ORF Name	in	Length         Length         Score         Probability           498         1497         370         5.4e-34           Locus         Name         Acc#           pir:S75887         S75887
Protein name  hypothetical prot  Description  ORF Name  4410135f256	in	Length         Length         Score         Probability           498         1497         370         5.4e-34           Locus         Name         Acc#           pir:S75887         S75887           NT         AA         Score         Probability           Length         Length         Probability
Protein name  hypothetical prot  Description  ORF Name	in	Length         Length         Score         Probability           498         1497         370         5.4e-34           Locus         Name         Acc#           pir:S75887         S75887           NT         AA         Score         Probability           Length         Length         Probability
Protein name  hypothetical prot  Description  ORF Name  4410135f256	in	Length         Length         Score         Probability           498         1497         370         5.4e-34           Locus         Name         Acc#           pir:S75887         S75887           NT         AA         Score         Probability           Length         Length         Probability

ORF Name 4486261 c1 121	NTID AAID Length Length Score Probability  194   5416   164   495
Protein name	Locus Name Acc#
Description	
NO-HIT	
ORF Name	NT AA Score Probability NTID AAID Length Length
4589627 <u>f3</u> 93	195 5417 199 600 257 5.1e-22
Protein name	Locus Name Acc#
ygge hypothetical	protein pir:H72114 H72114
Description	
ORF Name	$rac{ ext{NT}}{ ext{NTID}}$ $rac{ ext{AA}}{ ext{Length}}$ $rac{ ext{Ecore}}{ ext{Length}}$ $rac{ ext{Probability}}{ ext{Probability}}$
47157165 <u>_</u> c1 <u>_</u> 100	196 5418 75 228
Protein name	Locus Name Acc#
Description	
NO-HIT	
	NT AA
ORF Name	NTID AAID Length Length Probability
632012 <u></u> ±1 <u></u> 22	197 5419 82 249
Protein name	Locus Name Acc#
Description	
NO-HIT	
	NT AA Grand Duckahiliba
ORF Name	NTID AAID Length Length Score Probability
969687 <u></u> £1 <u>.</u> 28	<u>198</u> <u>5420</u> <u>140</u> <u>423</u>
Protein name	Locus Name Acc#
Description	
NO-HIT	

ORF Name 10631882_c2_238	<u>NTID</u>	<u>AAID</u> 5421	NT Length	AA Length	Score	Probability	
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability	
11148453 <u></u> 122	200	5422	127	384	126	3.9e-08	]
Protein name				Locus	Name	Acc#	
hypothetical protei	n yngA			pir:F69	892	F69892	
Description					1	- C	
							_
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
12381962 <u></u> £3 <u></u> 134	201	5423	1058	3177	2370	6.3e-246	1
Protein name			-	Locus	Name	Acc#	
hypothetical protei	n mexF			pir:T30	0830	T30830	
Description		1		· · · · · · · · · · · · · · · · · · ·			
					( )		
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
1256327 <u>_c3</u> _274	202	5424	149	50	96	1.6e-06	]
Protein name				Locus	Name	Acc#	
ct469 hypothetical	protein			pir:D72	060	D72060	
Description					1		
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
12690817 <u></u> .c2 <u></u> 225	203	5425	141 4	26	9. **		
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
<u> </u>			-				

ORF Name	NTID AAID	NT AA Score	Probability
12712827_£3_120	204 5426	353 1062 276	5.0e-24
Protein name		Locus Name	Acc#
conserved hypothe	etical protein	pir:F72386	F72386
Description			
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
12922202 <u>f3</u> 116	205 5427	316 951 132	3.0e-06
Protein name		Locus Name	Acc#
hypothetical pro	tein aq_380	pir:A70334	A70334
Description			
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
1370468 <u></u> .£3 <u></u> .118	206 5428	93 282 Name	Acc#
Protein name		Locus Name	ACC#
Description			1
NO-HIT			
ORF Name	NTID AAID	NT AA Score	Probability
138.78425 f2 72		134 405 368	8.9e-34
Protein name		Locus Name	Acc#
TIOCCIII IIIIII		sp:YYAH_BACSU	P37516
Description			
HYPOTHETICAL 14	.4 KD PROTEIN IN TETE	-EXOA INTERGENIC REGION	(ORFF)
	*	NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
14508425 <u></u> c1 <u></u> 192	208 5430	236 711 166	2.3e-12
Protein name		Locus Name	Acc#
hypothetical pro	otein MTH939	pir:G69225	G69225
Description			

ORF Name	NTID AAID	NT AA Score	Probability
14646956_c2_261	209 5431	82 249 61	0.023
Protein name		Locus Name	Acc#
mannanase		gp:U96771	U96771
Description			
Prevotella bryant mannanase genes, c	ii putative polygala omplete cds; and unk	cturonase,B-1,4-endogluc nowngenes.	anase, and
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	4
14705261_f1_15	210 5432	501 1506 115	0.00082
Protein name		Locus Name	Acc#
unknown protein	·	gp:BACCOMGA	
Description			
Bacillus subtilis operon, complete c		1,2,3,4,5,6,and 7) prote	ins incomG
	*	NT AA GGGTG	Probability
ORF Name	NTID AAID	Length Length Score	FIODADIIICY
14744010 <u></u> f3 <u>.</u> 129	211 5433	265 798 275	6.3e-24
Protein name		Locus Name	Acc#
conserved hypothe	tical protein yjkA	pir:E69851	E69851
Description	-	-	
		NT AA Coore	Probability
ORF Name	NTID AAID	Length Length	FIODADITICY
14876313 <u>_</u> £2 <u>_</u> 84	212 5434	667 2004 1443	1.1e-147
Protein name		Locus Name	Acc#
DNA ligase		gp:BST011676	AJ011676
DNA ligase Description	nermophilus lig gene		AJ011676

ORF Name	NTID AAID	NT AA Score	Probability
15641902_c3_288	213 5435	202 609 345	2.4e-31
Protein name		Locus Name	Acc#
conserved hypothe	etical protein AF2201	pir:A69525	A69525
Description		,	
		NT AA Grand	D
ORF Name	NTID AAID	Length Length Score	Probability
16204662 <u>f2</u> 65	214 <u>5436</u>	604 1815 134	3.1e-06
Protein name		Locus Name	Acc#
hypothetical pro	ein AF1867	pir:B69483	B69483
Description			
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	1
16828382 <u>f3</u> 13.7	215 5437	305 918 801	1.2e-79
Protein name		Locus Name	Acc#
		sp:PYRD_AQUAE	066461
Description		**	φ
(DHODEHASE)			
ORF Name	NTID AAID	NT AA Score	Probability
16832885 <u></u> c3 <u></u> 307	216 5438	431 1296 1724	1.8e-177
Protein name	* *	Locus Name	Acc#
hypothetical pro	tein	pir:JQ1020	JQ1020
Description			
		NT AA	Probability
ORF Name	NTID AAID	Length Length Score	
188752_f1_20		761 2286 380	3.5e-34
Protein name		Locus Name	ACC#
conserved hypoth	netical protein AF1878	pir:E69484	E69484
Description			

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
188905_c1_179	218 54	40	72 2	19		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
19645166_f3_136	219 54	441	345	L038	244	8.2e-20
Protein name			<u></u>	Locus	Name	Acc#
y .				sp:YQE	N_BACSU	P54459
Description						
HYPOTHETICAL 40.5	KD PROTEIN I	N COMEC	-RPST IN	TERGENIC	REGION	
111101111111						
			NT	AA	Score	Probability
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	
	NTID		<u>NT</u> Length	AA Length 2334	Score	2.5e-23
ORF Name  2051500_c3_266  Protein name	<u>NTID</u> 220 5	AAID 442	NT Length	AA Length 2334 Locu	Score 178 Name	2.5e-23 <u>Acc#</u>
ORF Name 2051500_c3_266	<u>NTID</u> 220 5	AAID 442	NT Length	AA Length 2334	Score 178 Name	2.5e-23
ORF Name  2051500_c3_266  Protein name	<u>NTID</u> 220 5	AAID 442	NT Length	AA Length 2334 Locu	Score 178 Name	2.5e-23 <u>Acc#</u>
ORF Name  2051500_c3_266  Protein name  conserved hypothe	<u>NTID</u> 220 5	AAID 442 aq_138	NT Length 777	AA Length  2334  Locu  pir:F7	Score 178 Name	2.5e-23 <u>Acc#</u>
ORF Name  2051500_c3_266  Protein name  conserved hypothe	<u>NTID</u> 220 5	AAID 442	NT Length 777	AA Length  2334  Locu  pir:F7  AA Length	178 s Name	2.5e-23 Acc# F70420
ORF Name  2051500_c3_266  Protein name  conserved hypother  Description	NTID  220  5  cical protein  NTID	AAID 442 aq_138	NT Length 777	AA Length  Locus  pir:F7  AA Length	Score  178  Name  0420  Score	Acc# F70420 Probability
ORF Name  2051500c3266  Protein name  conserved hypother  Description  ORF Name	NTID  220  5  cical protein  NTID	AAID  442  aq_1386	NT Length 777	AA Length  Locus  pir:F7  AA Length	178 s Name	2.5e-23 Acc# F70420
ORF Name  2051500c3266  Protein name  conserved hypother  Description  ORF Name  20741703t289	NTID  220  5  cical protein  NTID	AAID  442  aq_1386	NT Length 777	AA Length  Locus  pir:F7  AA Length	Score  178  Name  0420  Score	Acc# F70420 Probability

ORF Name 2151556_f1_27 Protein name	NTID AAID  [222] 5444	NT AA Score Length Length  Locus Name	Probability  Acc#
Description	_		
NO-HIT			
ORF Name	NTID AAID	NT AA Score	Probability
21541425 <u></u> c2 <u></u> 199	223 5445	186	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
21679062 <u>f1</u> 18	224 5446	283 852 206	1.3e-16
Protein name		Locus Name	Acc#
conserved hypothe	cical protein	pir:E72209	E72209
Description	-		
		NT AA G	
ORF Name	NTID AAID	Length Length Score	Probability
23.5.7.62f.122	225 5447	179 540	
Protein name		Locus Name	Acc#
Description			
NO-HIT	Ţ		
		NT AA Soore	1 1 1 1 2 1 2 1
ORF Name	NTID AAID	Length Length Score	Probability
23.6256.93 <u></u> t2 <u></u> 64	226 5448	265 798	
Protein name		Locus Name	Acc#
Description			
NO-HIT		× × × × × × × × × × × × × × × × × × ×	

23633312_f1_7 227 5449 248 747 193 3.1e-15	
Protein name Locus Name Acc#	
gp:APU72238 U72238	
Description	
Anabaena PCC7120 ORFR1, ORFR2, ORFR3, ORFR4, and ORFR5 genes, complete sequences.	
NT AA Grand Brobability	
ORF Name NTID AAID Length Length Score Probability	•
23635812_f1_9	
Protein name Locus Name Acc#	
conserved hypothetical protein AF1017 pir:A69377 A69377	
Description	
ORF Name NTID AAID Length Length Score Probability	-
23884561_f1_45229	
Protein name Locus Name Acc#	
63 kDa protein gp:MBU73653 U73653	
Description	
Description  Mycobacterium bovis 63 kDa protein, 47 kDa protein and clpB gene, complete cds.	4,
Mycobacterium bovis 63 kDa protein, 47 kDa protein and clpB gene, complete cds.	10.
Mycobacterium bovis 63 kDa protein, 47 kDa protein and clpB gene, complete	7
Mycobacterium bovis 63 kDa protein, 47 kDa protein and clpB gene, complete cds.  NT AAA Score Probability	7
Mycobacterium bovis 63 kDa protein, 47 kDa protein and clpB gene, complete cds.  NT AA Score Probability  ORF Name NTID AAID Length Length	7
Mycobacterium bovis 63 kDa protein, 47 kDa protein and clpB gene, complete cds.    NT AA   Score Probability   Page 10   Page	7

ORF Name	NTID AAID	NT AA Score	Probability
24316061_f1_6  Protein name	231 5453	Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Score	Probability
243343.93 <u>c1151</u>	. 232 5454	98 297	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Length Length	Probability
24412502 <u></u> c1 <u></u> 196	233 5455	94 285	
Protein name	* **	Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT <u>AA</u> Score Length Length	Probability
24485926 <u></u> £1 <u></u> 16	234 5456	476 1431	
Protein name		Locus Name	Acc#
Description			
NO-HIT			-
ORF Name	NTID AAID	NT AA Length Length Score	-
24494017 <u>c1</u> 169	235 5457	295 888 887	8.9e-89
Protein name		Locus Name	Acc#
hypothetical prote	in jhp0694	pir:F71901	F71901
Description			•

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24500032_£3_121	236	5458	612	1839	1501	7.7e-154
Protein name				Locus	s Name	Acc#
				sp:SYD	BACSU	032038
Description						
(ASPRS)						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24642760_c3_311	237	5459	415	1245	1990	1.2e-205
Protein name				Locus	s Name	Acc#
L-fucose isomerase				gp:AF1	37263	AF137263
Description						
Bacteroides thetaio gene cluster, and RN complete cds.			_		_	L L
			2700	73.73		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 24726550t261	NTID 238	<u>AAID</u> 5460	Length	-	Score	Probability
			Length	Length	Score S Name	Probability  Acc#
24726550 <u></u> t2 <u></u> 61			Length	Length	-	. *
24726550 <u></u> t2 <u></u> 61 Protein name			Length	Length	-	. *
Protein name  Description			Length	Length 411 Locus	3 Name	Acc#
Protein name  Description			Length	Length	-	. *
Protein name  Description  NO-HIT	238	5460	Length  136  NT  Length	Length Locus	3 Name	Acc#
24726550_t2_61Protein name  Description  NO-HIT  ORF Name	NTID	5460	Length  136  NT  Length	Length Locus  AA Length	S Name	Acc# Probability
24726550t261	NTID 239	5460	Length  136  NT  Length	Length Locus  AA Length	Score  88 Name	Acc# Probability 0.013

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24897943_f1_12  Protein name	240	5462	61 1	Locus	s Name	Acc#
Description NO-HIT						
110-1111	·					
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
25316.92t121	241	5463	105	318	217	8.9e-18
Protein name				Locus	Name	Acc#
				gp:AB02	24563	AB024563
Description					т.	
Bacillus halodurans	gene for	YFIL, YF	IM, YFIN	, YHDE, I	MP and .	ARGE, complete
					<del></del>	
			NT	AA		Descharation of the same
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 2617125fl40	NTID	AAID 5464	Length		Score	Probability
-			Length	Length	Score S Name	Probability  Acc#
			Length	Length		*
2617125f140 Protein name			Length	Length		*
Protein name  Description			Length	Length  Locus		Acc#
Protein name  Description			Length	Length		*
Protein name  Description  NO-HIT	NTID	5464	Length  NT Length	Length  Locus	s Name	Acc#
Protein name  Description  NO-HIT  ORF Name	NTID	AAID	Length  NT Length	Length Locus  AA Length	Score	Acc# Probability
Protein name  Description  NO-HIT  ORF Name  263.76.540	NTID 243  on system	AAID 5465	Length  NT Length  255	Length Locus  AA Length	Score 106 Name	Acc# Probability 0.0015

ORF Name 26578375 f3 100	NTID	AAID	NT Length	AA Length	Score	Probability
Protein name	] []			J	Name	Acc#
Description NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2741543_f3_111	. 245	5467	608 1	.827	111	0.0053
Protein name				Locus	Name	Acc#
				sp:SECY	_ANTSP	Q37143
Description						
PREPROTEIN TRANSLO	CASE SECY S	UBUNIT			*	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2851577 <u></u> £1 <u>4</u> 1	246	5468	297 8	94	296	3.8e-26
Protein name				Locus	Name	Acc#
Protein name XylR				Locus gp:BSU1		<u>Acc#</u> U15985
XyIR	rmophilus e	ndo-beta-	1,4-xyla	gp:BSU1	5985	U15985
XylR  Description  Bacillus stearother	rmophilus e	ndo-beta-	1,4-xyla	gp:BSU1	5985 nA) gene	U15985
XylR  Description  Bacillus stearother	rmophilus en	ndo-beta- AAID		gp:BSU1	5985	U15985
Description  Bacillus stearother  cds.	NTID		NT Length	gp:BSU1 nase (xy	5985 nA) gene	U15985
Description  Bacillus stearother  cds.  ORF Name	NTID	AAID	NT Length	gp:BSU1 nase (xy	5985 nA) gene Score 433	U15985 ,complete Probability
Description  Bacillus stearother  cds.  ORF Name	NTID	AAID	NT Length	gp:BSU1 nase (xy) AA Length	5985 nA) gene Score 433 Name	U15985 ,complete  Probability  1.1e-40
Description  Bacillus stearother  cds.  ORF Name	NTID	AAID	NT Length	gp:BSU1  nase (xy)  AA  Length  Locus	5985 nA) gene Score 433 Name	U15985  , complete  Probability  1.1e-40  Acc#

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
29537532_±3_117	248	5470	101	306	113	9.3e-07	
Protein name				Locu	s Name	Acc#	
hypothetical prot	ein Rv2816c			pir:C7	0691	C70691	
Description							
			).TIT	73.73			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
33205013 <u>_</u> c1 <u>_</u> 157	249	5471	247	744	605	6.8e-59	
Protein name				Locu	s Name	Acc#	
				sp:TRM	D_BACSU	031741	
Description				<u> </u>			
METHYLTRANSFERASE	Ξ)						
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
34015677 <u></u> £3 <u></u> 131	250	5472	279 8	340	218	7.0e-18	7
Protein name			L	Locus	Name	Acc#	
			121	Locus		Acc# A71519	
Protein name							
Protein name  hypothetical prot				pir:A7			
Protein name  hypothetical prot		AAID	NT Length				
Protein name  hypothetical prot  Description	ein CT398	<u>AAID</u>	<u>Length</u>	pir:A7	1519	A71519	
Protein name  hypothetical prot  Description  ORF Name	ein CT398		<u>Length</u>	Pir:A7	Score	A71519  Probability	
Protein name  hypothetical prot  Description  ORF Name  34068760_f1_28	NTID 251	5473	<u>Length</u>	Pir:A7	Score  [599  Name	A71519  Probability  2.9e-58	
Protein name  hypothetical prot  Description  ORF Name  34068760_f1_28  Protein name	NTID 251	5473	<u>Length</u>	AA Length	Score  [599  Name	A71519  Probability  2.9e-58  Acc#	
Protein name  hypothetical prot  Description  ORF Name  34068760_f1_28  Protein name  conserved hypothe	NTID 251	5473	Length 452	AA Length .359 Locus	Score  [599  Name	A71519  Probability  2.9e-58  Acc#	
Protein name  hypothetical prot  Description  ORF Name  34068760_f1_28  Protein name  conserved hypothe	NTID 251	5473	<u>Length</u>	AA Length	Score  [599  Name	A71519  Probability  2.9e-58  Acc#	
Protein name  hypothetical prot  Description  ORF Name  34068760_f1_28  Protein name  conserved hypothe  Description	NTID  Control  Contro	n yqfO	Length  452 1  NT Length	AA Length Locus pir:A63	Score [599 S Name	A71519  Probability  2.9e-58  Acc#  A69954	
Protein name  hypothetical prot  Description  ORF Name  34068760_f1_28  Protein name  conserved hypothe  Description  ORF Name	NTID  Control  Contro	n yqfO	Length  452 1  NT Length	AA Length  Locus  pir:A69	Score [599 S Name	A71519  Probability  2.9e-58  Acc#  A69954	
Protein name  hypothetical prot  Description  ORF Name  34068760_f1_28  Protein name  conserved hypothe  Description  ORF Name  35345057_c2_240	NTID  Control  Contro	n yqfO	Length  452 1  NT Length	AA Length  Locus  pir:A69	Score  599 S Name 9954 Score	Probability  2.9e-58  Acc#  A69954  Probability	
Protein name  hypothetical prot  Description  ORF Name  34068760_f1_28  Protein name  conserved hypothe  Description  ORF Name  35345057_c2_240  Protein name	NTID  Control  Contro	n yqfO	Length  452 1  NT Length	AA Length  Locus  pir:A69	Score  599 S Name 9954 Score	Probability  2.9e-58  Acc#  A69954  Probability	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
35432303_c2_228	253	5475	396 1	191	778	3.2e-77
Protein name				Locus	Name	Acc#
hypothetical protei	n HP0049			pir:A64	526	A64526
Description						
			NT	AA	G	Dwobobility
ORF Name	NTID	AAID	Length	Length	Score	Probability
35832062 <u></u> ±3 <u></u> 108	254	5476	815	2448	173	2.9e-09
Protein name				Locus	Name	Acc#
				sp:YBJ	_ECOLI	P75831
Description						
HYPOTHETICAL ABC TE	RANSPORTER	ATP-BIND	ING PROTE	EIN YBJZ		
			NT	AA		Dechability
ORF Name	NTID	AAID	<u>NT</u> Length		Score	Probability
ORF Name 36362551_c2_232		<u>AAID</u> 5477	Length	<u>AA</u>	Score 801	Probability
			Length	AA Length		
36362551 <u></u> c2 <u></u> 232			Length	AA Length	801	1.2e-79
36362551 <u></u> c2 <u></u> 232			Length	AA Length	801 S Name	1.2e-79 <u>Acc</u> #
36362551_c2_232 Protein name			Length	AA Length	801 S Name	1.2e-79 <u>Acc</u> #
Protein name  Description			Length	AA Length Locus sp:BIO	801 S Name F BACSH	1.2e-79 <u>Acc#</u> P22806
Protein name  Description			Length	AA Length	801 S Name	1.2e-79 <u>Acc</u> #
Protein name  Description  LIGASE)	NTID	5477	Length  396  NT Length	AA Length Locus sp:BIO	801 S Name F BACSH	1.2e-79 <u>Acc#</u> P22806
Protein name  Description  LIGASE)  ORF Name	NTID	5477 AAID	Length  396  NT Length	AA Length Locus Sp:BIO	801 S Name F BACSH	1.2e-79 <u>Acc#</u> P22806
Protein name  Description  LIGASE)  ORF Name  3906625_c3_309	NTID	5477 AAID	Length  396  NT Length	AA Length Locus Sp:BIO	801 S Name F_BACSH Score	1.2e-79  Acc# P22806  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
3923465_f3_135	257	5479	261 7	86	510	7.9e-49	]
Protein name				Locus	Name	Acc#	
amp nucleosidase				pir:A72	021	A72021	
Description							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
3928336 <u></u> c2262	258	5480	464 1	.395			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			ידידע	λλ			_
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	_
ORF Name 3.93.27.56t274	<u>NTID</u>	<u>AAID</u>	Length		Score 518	Probability	]
			Length	Length			]
3.93.2.7.5.6t27.4			Length	Length	518 Name	1.1e-49	
3932756t274 Protein name			Length	Length 416 Locus	518 Name	1.1e-49 Acc#	
3932756t274 Protein name  OprM	259	5481	Length 471	Length 416  Locus gp:AB01	518 Name	1.1e-49 Acc#	
Protein name OprM Description	259	5481	Length  471 [1	Length Locus gp:AB01	518 Name	1.1e-49 Acc# AB011381	] ;
Protein name OprM Description	259	5481	Length 471	Length 416  Locus gp:AB01	518 Name	1.1e-49 Acc#	]
3932756t2_74  Protein name  OprM  Description  Pseudomonas aerugin	osa gene	for OprM,	Length  471 I  complete  NT Length	Length Locus gp:AB01 c cds.	518 : Name -1381	1.1e-49 Acc# AB011381	<u></u>
3932756_t2_74  Protein name OprM Description Pseudomonas aerugin ORF Name	osa gene	for OprM,	Length  471 I  complete  NT  Length	Length Locus gp:AB01 c cds. AA Length	518 : Name -1381	1.1e-49 Acc# AB011381	]
3932756_t2_74	osa gene	for OprM,	Length  471 I  complete  NT  Length	Length Locus gp:AB01 c cds. AA Length	518 S Name 1381 Score	Acc# AB011381  Probability	3

ORF Name NTID AAID Length Length Probability	<u>r</u>
3946962_c1_163	
Protein name Locus Name Acc#	
Description	
NO-HIT	
ORF Name NTID AAID Length Length Probability	<u>-</u>
3.98.56.4.0c226.0	
Protein name Locus Name Acc#	
Description	
NO-HIT	
ORF Name NTID AAID Length Length Score Probability	- 0
4103383 <u>c2</u> 201 263 5485 125 378 164 3.7e-12	
Protein name Locus Name Acc#	
sp:YBDF_ECOLI	
Description	
HYPOTHETICAL 14.1 KD PROTEIN IN NFNB-ENTD INTERGENIC REGION	7
ORF Name NTID AAID Length Length Score Probability	
41307 <u></u> £1 <u></u> 14	
Protein name Locus Name Acc#	
sp:Y978_METJA Q58388	
Description	
HYPOTHETICAL PROTEIN MJ0978	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4343942_£2_75	265	5487	383	152	443	1.0e-41
Protein name				Locus	Name	Acc#
hypothetical prote	ein mexE	-		pir:T30	)829	T30829
Description						
			NT	AA	G	Dechability
ORF Name	NTID	AAID	Length	Length	Score	Probability
4531536 <u></u> ±3 <u></u> 112	266	5488	412	L239	150	7.0e-10
Protein name				Locus	s Name	Acc#
				gp:YP1	02KB	AL031866
Description						
Yersinia pestis 1	02 kbases u	nstable r	egion: fi	rom 1 to	119443.	
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 4535677c1154	. 1	<u>AAID</u> 5489	Length		<u>Score</u> 241	Probability 2.5e-20
×	. 1		Length	Length		
45356.7.7 <u></u> c1 <u></u> 154	. 1		Length	Length  Locu	241	2.5e-20
45356.7.7 <u></u> c1 <u></u> 154	. 1		Length	Length  Locu	241 s Name	2.5e-20 <u>Acc#</u>
45356.7.7c1154 Protein name	267	5489	Length 211	Length  Locu	241 s Name	2.5e-20 <u>Acc#</u>
Protein name  Description	267	5489	Length 211	Length Locu  Sp:CAT	241 s Name	2.5e-20 <u>Acc#</u>
Protein name  Description  CHLORAMPHENICOL A	267	5489	Length 211	Length  Locu	241 s Name	2.5e-20 <u>Acc#</u>
Protein name  Description  CHLORAMPHENICOL A  ORF Name	CETYLTRANSF	5489 ERASE III	Length  211  NT Length	Length Locu Sp:CAT	241 s Name 3_ECOLI	2.5e-20 <u>Acc#</u> P00484
Protein name  Description  CHLORAMPHENICOL A  ORF Name  46.910.32_c2_23.3	CETYLTRANSF	5489 ERASE III	Length  211  NT Length	Length Locus Sp:CAT  AA Length	241 s Name 3_ECOLI	2.5e-20 <u>Acc#</u> P00484
Protein name  Description  CHLORAMPHENICOL A  ORF Name  46.910.32_c2_2.33	CETYLTRANSF	5489 ERASE III	Length  211  NT Length	Length Locus Sp:CAT  AA Length	241 s Name 3_ECOLI Score	2.5e-20 Acc# P00484 Probability
Protein name  Description  CHLORAMPHENICOL A  ORF Name  46.910.32_c2_23.3	CETYLTRANSF	5489 ERASE III	Length  211  NT Length	Length Locus Sp:CAT  AA Length	241 s Name 3_ECOLI Score	2.5e-20 Acc# P00484 Probability

		$\underline{ ext{NT}}$	AA Score	Probability
ORF Name	NTID AAII	<u>Length</u>	Length —	
4691061_c3_294	269 5491	94 2	72	0.020
Protein name			Locus Name	Acc#
NADH dehydrogenase	subunit 4L		gp:BMMITOCH01	AF110610
Description				
Boophilus micropl NADH dehydrogenase tRNA-Pro genes, co gene, partial cds,	subunit 4L (ND4L)	) gene, com <u>r</u> and NADHdehy	oletecas; tRNA-1. /drogenase subun	it 6 (ND6)
	NTID AAI	<u>NT</u> D Length	AA Length Score	Probability
ORF Name			120	
4897087_£1_19	270 5492			Acc#
<u>Protein name</u>			Locus Name	<u>ACC#</u>
Description				
NO-HIT				
NO-HIT	NTID AAI	<u>NT</u> D Length	AA Score	Probability
		D <u>Length</u>	— Score	Probability  8.6e-06
ORF Name		D <u>Length</u>	Length Score	
ORF Name 4898537t111		D <u>Length</u>	Length Score	8.6e-06
ORF Name 4898537t111		D <u>Length</u>	Length Score  Length 136  Locus Name	8.6e-06 <u>Acc#</u>
ORF Name 48.9853.7t111	271   5493	D <u>Length</u>	Length Score  Length 136  Locus Name	8.6e-06 <u>Acc#</u>
ORF Name  48.9853.7_f1_11  Protein name  Description	271   5493	D <u>Length</u>	Length Score  Locus Name  Sp:Y797_METJA	8.6e-06 <u>Acc#</u> Q58207
ORF Name  48.9853.7_f1_11  Protein name  Description	271   5493	D Length  749	Length Score  Locus Name  Sp:Y797_METJA	8.6e-06 <u>Acc#</u>
ORF Name  489853.7fl11  Protein name  Description  HYPOTHETICAL PROT	271 5493 EIN MJ0797 NTID AAI	D Length  749  NT Length	Length Score  Locus Name  Sp:Y797_METJA  AA Score	8.6e-06 <u>Acc#</u> Q58207
ORF Name  489853.7fl11  Protein name  Description  HYPOTHETICAL PROT	271 5493 EIN MJ0797 NTID AAI	D Length  749  NT Length	Length Score  Locus Name  Sp:Y797_METJA  AA Length Score	8.6e-06  Acc# Q58207  Probability
ORF Name  48.98.53.7t111  Protein name  Description  HYPOTHETICAL PROT  ORF Name  49.73.76.5t310.4	271 5493 EIN MJ0797 NTID AAI	D Length  749  NT D Length  141	Length Score  Locus Name  Sp:Y797_METJA  AA Length Score  128	8.6e-06 Acc# Q58207  Probability  1.6e-07

			$\underline{ ext{NT}}$	AA	Score	Probability
ORF Name	$\overline{ ext{NTID}}$	AAID L	Length	Length	<u>pcore</u>	Probability
5084381_c3_310	273 54	95 3	331 99	96	1166	2.4e-118
Protein name				Locus	Name	Acc#
FucR				gp:AF13	7263	AF137263
Description						**
Bacteroides thete gene cluster, and complete cds.						
ORF Name	NTID	AAID L	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
5187812_f1_13	274 54	96 2	28 68	37	596	6.1e-58
Protein name				Locus	Name	Acc#
				sp:YF08	METJA	Q58903
Description					**	
HYPOTHETICAL ABC	TRANSPORTER AT	P-BINDING	G PROTEI	N MJ150	8	
HYPOTHETICAL ABC	TRANSPORTER A	P-BINDING			·	
HYPOTHETICAL ABC	TRANSPORTER AT		NT	AA Length	Score	Probability
	NTID	AAID L	NT	AA Length	·	Probability 7.5e-21
ORF Name	NTID	AAID L	NT Length	AA Length	Score	
ORF Name 5191942_f3_128	<u>NTID</u> 275   54	<u>AAID</u> <u>L</u>	NT Length	AA Length	Score 246 Name	7.5e-21
ORF Name  5191942_f3_128  Protein name  amino acid ABC ti	<u>NTID</u> 275   54	<u>AAID</u> <u>L</u>	NT Length	AA Length	Score 246 Name	7.5e-21 <u>Acc#</u>
ORF Name  5191942_f3_128  Protein name  amino acid ABC tr	<u>NTID</u> 275   54	<u>AAID</u> <u>L</u>	NT Length	AA Length 30 Locus pir:H72	Score 246 Name	7.5e-21 <u>Acc#</u>
ORF Name  5191942_f3_128  Protein name  amino acid ABC tr	<u>NTID</u> 275   54	AAID L	NT Gength 63	AA Length	Score 246 Name	7.5e-21 <u>Acc#</u>
ORF Name  5191942_f3_128  Protein name  amino acid ABC transportein  Description	NTID  275 54  cansporter, ATE	AAID L	NT Gength 63	AA Length Locus pir:H72  AA Length	<u>Score</u> 246 <u>Name</u> 356	7.5e-21 Acc# H72356
ORF Name  5191942_f3_128  Protein name  amino acid ABC transprotein  Description  ORF Name	NTID  275 54  cansporter, ATE	AAID L	NT Jength 109 63 NT Jength	AA Length Locus pir:H72  AA Length	Score  246  Name  356  Score	7.5e-21  Acc#  H72356  Probability
ORF Name  5191942_f3_128  Protein name  amino acid ABC tr protein  Description  ORF Name  547082_f1_36	NTID  275 54  cansporter, ATE	AAID L	NT Jength 109 63 NT Jength	AA Length Locus pir:H72  AA Length	Score  246  Name  356  Score  93  Name	7.5e-21 Acc# H72356  Probability 0.00016
ORF Name  5191942_f3_128  Protein name  amino acid ABC transportein  Description  ORF Name  547082_f1_36  Protein name	NTID  275 54  cansporter, ATE	AAID L	NT Jength 109 63 NT Jength	AA Length  Locus  Pir:H72  AA Length  .3 Locus	Score  246  Name  356  Score  93  Name	7.5e-21 Acc# H72356  Probability 0.00016

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
5859625_c3_299	277	5499	62	189		
Protein name			L	Locus	. Name	Acc#
Description						<del></del>
NO-HIT						
INO-HII						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
6412812 <u></u> c2198	278	5500	686	2061	2246	8.7e-233
Protein name				Locus	Name	Acc#
high temperature pro	tein HtpG	j		gp:AF17	6245	AF176245
Description						
Porphyromonas gingiveds.	alis high	ı temperat	ure prot	ein HtpG	(htpG)	gene, complete
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	<u>BCOLE</u>	FIODADITICY
6.7.23.262£256	279	5501	133	102		
Protein name				Locus	Name	Acc#
Description						
NO-HIT					*	
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
78.7567 <u>_</u> £3 <u>_</u> 139	280	5502	303	912	659	1.3e-64
Protein name				Locus	Name	Acc#
dihydrodipicolinate	synthase			pir:B72	246	B72246
Description						

			NT	AA Tanath	Score	Probability
ORF Name	NTID	AAID	Length	Length	FARR	7  1.1e-193
907974_c1_149	281	5503	745 2	238	1877	
Protein name				Locus	Name	Acc#
				sp:MECE	B_BACSU	P37571
Description						
NEGATIVE REGULATOR	OF GENETI	C COMPETE	NCE MECB			
			NT	AA		Drobability
ORF Name	NTID	AAID	Length	Length	Score	Probability
9851505_f3_113	282	5504	238 7	17	106	1.5e-05
Protein name				Locus	Name	Acc#
hypothetical protei	n		-	gp:SEL2	243707	AJ243707
Description						
Synechococcus elong	jatus petB	gene, pe	tD gene a	and ORF1	· *	* * *
1						
			NTT			
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
ORF Name 10969062_t3_151		<u>AAID</u>  5505	Length	AA	Score	Probability 6.1e-74
			Length	<u>AA</u> Length 771		1
10969062 <u></u> £3. <u></u> 151	283	5505	Length	<u>AA</u> Length 771	747 s Name	6.1e-74
10969062_t3_151 Protein name	283	5505	Length	AA Length 771 Locu	747 s Name	6.1e-74 Acc#
Protein name  ATP synthase F1, su	283	5505	Length	AA Length 771 Locus	747 s Name	6.1e-74 Acc#
Protein name  ATP synthase F1, su	283	5505	Length	AA Length 771 Locu	747 s Name	6.1e-74 Acc#
Protein name  ATP synthase F1, su  Description  ORF Name	283  abunit alp	5505 na	Length  256  NT  Length	AA Length 771 Locus pir:F7	747 s Name 2231	6.1e-74 Acc# F72231
Protein name  ATP synthase F1, su  Description  ORF Name	283  abunit alp	5505 ha <u>AAID</u>	Length  256  NT  Length	AA Length Locus pir:F7	747 S Name 2231 Score	6.1e-74  Acc#  F72231  Probability
Protein name  ATP synthase F1, su  Description  ORF Name	283 abunit alp  NTID 284	5505 ha <u>AAID</u>	Length  256  NT  Length	AA Length Locus pir:F7	747 s Name 2231 Score 565 s Name	6.1e-74  Acc#  F72231  Probability  1.2e-54
Protein name  ATP synthase F1, su  Description  ORF Name  133562_t3_116	283 abunit alp  NTID 284	5505 ha <u>AAID</u>	Length  256  NT  Length	AA Length  Locus  Pir:F7  AA Length  798  Locus	747 s Name 2231 Score 565 s Name	Acc# F72231  Probability  1.2e-54  Acc#

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13672255_f2_95  Protein name	285	5507	146	Locus	s Name	Acc#
Description						
NO-HIT						
	*		NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length		
14501312 <u></u> c3 <u></u> 311	286	5508	461	L386	473	6.6e-45
Protein name				_	Name	Acc#
conserved hypotheti protein HP1184	cal integr	ral membra	ane	pir:H64	1667	H64667
Description				j		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
14511007 <u></u> f1 <u></u> 2	287	5509	330	993		
Protein name				Locus	Name	Acc#
Description						
NO-HIT					<u> </u>	
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
14645437 <u></u> ±3 <u></u> 135	288	5510	768	2307	117	0.0015
Protein name		<u></u>		Locus	Name	Acc#
conserved hypotheti	cal protei	in yknZ		pir:E69	858	E69858
conserved hypotheti Description	cal protei	in yknZ		pir:E69	858	E69858
	cal prote	in yknZ	NI	-	858	E69858
	cal protei	in yknZ	<u>NT</u> Length	Pir:E69	Score	E69858  Probability
Description			Length	AA		
Description ORF Name	NTID	AAID	Length	AA Length	Score	Probability
Description  ORF Name  14969692_f1_37	NTID 289	<u>AAID</u> 5511	Length 404	AA Length	Score  232  Name	Probability 4.9e-17

ORF Name	NTID AAID	NT AA Length Length	Score Probability
15104137_f1_1	290  5512	466   1401	
Protein name		Locus	Name Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score Probability
15673443c3323	291 5513	1395	1.0e-78
Protein name		Locus	Name Acc#
Salmonella typhimu	rium transcriptiona	gp:STYS1	MF1 AF170176
Description			
Salmonella typhimu	rium fragment STMF1	•	
		NT AA	
ORF Name	NTID AAID	Length Length	Score Probability
158187f3121	. 292 5514	409 1230	988 1.8e-99
Protein name		Locus	Name Acc#
		sp:URAA_	HAEIN P45117
Description			
PROBABLE URACIL PE	RMEASE (URACIL TRANS	SPORTER)	
		<del></del>	
		No. 7.7	
ORF Name	NTID AAID	NT AA Length Length	Score Probability
ORF Name 16172682_f1_45	NTID AAID	Length Length	Score Probability 3.4e-78
		Length Length	3.4e-78
16172682_f1_45		Length Length 310 933	3.4e-78  Name Acc#
16172682_f1_45		Length Length  310 933 E	3.4e-78  Name Acc#
16172682t145 Protein name	293 5515	Length Length  310 933 E	3.4e-78  Name Acc#

·		NT AA Grana Brobability
ORF Name	NTID AAID	Length Length Score Probability
16828462_f1_46	294 5516	292 879 470 1.4e-44
Protein name		Locus Name Acc#
		sp:ATPG_BACSU P37810
Description		
ATP SYNTHASE GAMMA	CHAIN,	
		NT AA Grana Brobability
ORF Name	NTID AAID	Length Length
19945317_c2_264	295 5517	285 858 120 2.0e-05
Protein name		Locus Name Acc#
3',5'-cyclic-nucle cpdA homolog MTH178	otide phosphodiest :Icc related prote	erase, pir:F69104 F69104 in
Description		
		**
ODE Name	NTID AAID	NT AA Score Probability Length Length
ORF Name 21990930cl217		7 326 981
		Locus Name Acc#
<u>Protein name</u>		<u> </u>
Description		
NO-HIT		
		NT AA Grave Probability
ORF Name	NTID AAID	NT AA Score Probability Length Length
22464182 <u></u> c1 <u></u> 215	297 5519	328 987
Protein name		Locus Name Acc#
Description		
NO-HIT	*	

	A service of the serv	****	
ORF Name	NTID AAID	NT AA Length Length Score	Probability
2346936_f3_153	298 5520	463 1392 153	8.5e-08
Protein name		Locus Name	Acc#
HelC		gp:LPU11704	U11704
Description			
Legionella pneum	ophila HelC (helC) ge	ene, complete cds.	,
ORF Name	NTID AAID	NT AA Length Length Score	Probability
23694686_£1_35	299 5521	225 678 179	9.5e-14
Protein name		Locus Name	Acc#
		sp:GS1_HUMAN	Q08623
Description		*	-
GS1 PROTEIN			
ORF Name	NTID AAID	NT AA Length Length Score	Probability
23829510 <u></u> .c3 <u></u> 325	300 5522	228 687 283	9.0e-25
Protein name		Locus Name	Acc#
transcription reg	gulator, crp family	pir:F72285	F72285
Description			
ORF Name	NTID AAID	NT AA Length Length Score	Probability
24009530 <u></u> c1 <u></u> 182	301 5523	70 213	
Protein name		Locus Name	Acc#
Description			
NO-HIT			

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24010962_f1_25	302	5524	131	396			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							]
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
241113.75£3110	. 303	5525	68 2	207			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							]
			2777	77			
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
ORF Name 24219032_t2_97		<u>AAID</u>	Length		Score 172	Probability 5.2e-13	
			Length	Length			
24219032 <u>t2</u> 97			Length	Length	172 Name	5.2e-13	
24219032 <u>t2</u> 97			Length	Length 279 Locus	172 Name	5.2e-13 <u>Acc#</u>	
24219032f297 Protein name	. 304	5526	Length	Length 279 Locus	172 Name	5.2e-13 <u>Acc#</u>	
Protein name  Description	. 304	5526	Length  92  2  PROTEIN)	Length 279  Locus sp:ATPI	172 Name	5.2e-13 <u>Acc#</u>	× ×
Protein name  Description	. 304	5526	Length	Length 279 Locus	172 Name	5.2e-13 <u>Acc#</u>	
Protein name  Description  ATP SYNTHASE C CHA	IN, (LIPID	5526	Length  92  PROTEIN)  NT  Length	Length Locus Sp:ATPI	172 Name	5.2e-13 Acc# P12409	
Protein name  Description  ATP SYNTHASE C CHA	IN, (LIPID	-BINDING	Length  92  PROTEIN)  NT  Length	Length Locus Sp:ATPI  AA Length	Name  ANASP  Score	5.2e-13 Acc# P12409 Probability	
Protein name  Description  ATP SYNTHASE C CHA  ORF Name  24254417_f2_53	NTID 305  Don system	BINDING  AAID  5527  regulatory	Length  92  PROTEIN)  NT Length  379  1	Length Locus Sp:ATPI  AA Length	I72  Name  ANASP  Score  480  Name	5.2e-13 Acc# P12409  Probability  1.2e-45	

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
24333127_c3_332	306	5528	532	1599	748	1.3e-76
Protein name				Locus	Name	Acc#
				sp:YIEN	_ECOLI	P31473
Description				<u></u>		
HYPOTHETICAL 56.4	KD PROTEIN	IN ASNA-	KUP INTE	RGENIC RE	GION	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24398417_c2_280	307	5529	996	2991	108	3.4e-15
Protein name				Locus	Name	Acc#
hypothetical protei	n jhp0336			pir:C71	944	C71944
Description				·		
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24399035c3312	308	5530	60	183	44	0.049
Protein name	,			Locus	Name	Acc#
nonstructural prote	ein			gp:AF01	.2732	AF012732
Description				· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	
Bovine viral diarrh	nea virus	strain Ya	k nonstru	ictural p	rotein (	(p125) mRNA,
partial cds.						
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
24408437 <u></u> c2 <u></u> 228	309	5531	122	369	265	7.3e-23
Protein name				Locus	Name	Acc#
				sp:THIO	BORBU	051088
Description						
THIOREDOXIN (TRX)						

	NT AA Grana Doobabilitus
ORF Name	NTID AAID Length Length Score Probability
24414153_±1_20	310 5532 448 1347 83 0.0040
Protein name	Locus Name Acc#
unknown	gp:U96771 U96771
Description	
	ii putative polygalacturonase,B-1,4-endoglucanase, and omplete cds; and unknowngenes.
	NT AA C
ORF Name	NTID AAID Length Length Score Probability
24489452_c2_269	311 5533 560 1683 526 2.0e-58
Protein name	Locus Name Acc#
long-chain-fatty-	acid CoA ligase pir:D70386 D70386
Description	
ORF Name	NTI AA Score Probability NTID AAID Length Length
24644068 <u></u> t2 <u></u> 94	312    5534    83    252    158    1.6e-11
Protein name	Locus Name Acc#
	sp:ATPE_CHLLI P35111
Description	
ATP SYNTHASE EPS	LON CHAIN,
ATP SYNTHASE EPSI	
ORF Name	NT AA  NTID AAID Length Length Score Probability
	NTID AAID Length Length Score Probability
ORF Name	NTID AAID Length Length Score Probability
ORF Name 24808312t149	NTID AAID Length Length Score Probability  313   5535   189   567   308   3.8e-26
ORF Name 24808312t149	NTID AAID Length Length Score Probability  1313   5535   189   567   308   3.8e-26  Locus Name Acc#
ORF Name  24808312t149  Protein name	NTID AAID Length Length Score Probability  1313   5535   189   567   308   3.8e-26  Locus Name Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24875042_f3_147	314	5536	76	231	27	7 11	
Protein name				Locus	s <u>Name</u>	Acc#	
Description NO-HIT							
NO-HII							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24881577 <u></u> £2 <u></u> 85	. 315	5537	248	747	210	5.1e-16	
Protein name				Locus	Name	Acc#	
				sp:XYN	B_BUTFI	P26223	
Description							
D-XYLAN XYLANOHYDR	OLASE B)						
			NT	AA		*	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 25479842_c3_315		<u>AAID</u>	Length		Score	Probability	
			Length	Length	Score  S Name	Probability  Acc#	4
25479842c3315			Length	Length			3
25479842_c3_315 Protein name			Length	Length			
Protein name  Description			Length	Length 204 Locus		Acc#	
Protein name  Description			Length	Length			
Protein name  Description  NO-HIT	316 NTID	5538	Length  67  NT  Length	Length  Locus	s Name	Acc#	
Protein name  Description  NO-HIT  ORF Name	316 NTID	5538 AAID	Length  67  NT  Length	Length Locus  AA Length	Score	Acc# Probability	
25479842_c3_315	NTID	5538 AAID	Length  67  NT  Length	Length Locus  AA Length	Score 129 S Name	Acc# Probability  3.7e-05	
Protein name  Description  NO-HIT  ORF Name  26446928_f3_113  Protein name	NTID	5538 AAID	Length  67  NT  Length	Length  Locus  AA  Length  765  Locus	Score 129 S Name	Acc#  Probability  3.7e-05  Acc#	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
26831386_f3_139	318	5540	349	1050	507	1.7e-48
Protein name				Locus	Name	Acc#
				sp:PYRI	_ECOLI	P05021
Description						
(DHODEHASE)						
ORF Name	NTID	DIAA	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2942827_c3_342	319	5541	1057	31.74	214	3.5e-28
Protein name				Locus	s Name	Acc#
probable ATP-depe	ndent helic	ase		pir:A7	1805	A71805
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 29494000		<u>AAID</u> 5542	Length		Score	Probability 4.8e-27
			Length	Length		
29494000 <u></u> ±1 <u></u> 48			Length	Length  1095  Locu	309	4.8e-27
29494000 <u></u> ±1 <u></u> 48			Length	Length  1095  Locu	309 s Name	4.8e-27 Acc#
29494000f148 Protein name	320	5542	Length	Length  1095  Locu	309 s Name	4.8e-27 Acc#
Protein name  Description	320	5542	Length	Length  1095  Locu  sp:CZC	309 s Name	4.8e-27 <u>Acc#</u> P94176
Protein name  Description	320	5542	Length	Length  1095  Locu	309 s Name	4.8e-27 Acc#
Protein name  Description  CATION EFFLUX SYS	320 STEM PROTEIN	5542 N CZCB	Length  NT Length	Length  1095  Locu  Sp:CZC	309 s Name B_ALCSP	4.8e-27 <u>Acc#</u> P94176
Protein name  Description  CATION EFFLUX SYS	320 STEM PROTEIN	5542 N CZCB	Length  364  NT Length	Length  Locu  Sp:CZC  AA  Length	309 s Name B_ALCSP Score	Acc# P94176  Probability
Protein name  Description  CATION EFFLUX SYS  ORF Name	320 STEM PROTEIN NTID	5542 N CZCB  AAID  5543	Length  NT Length  208	Length  Locu  Sp:CZC  AA  Length	309 s Name B_ALCSP  Score 213 s Name	4.8e-27 Acc# P94176  Probability  5.3e-16

ODE Name	MITTE	2 <b>7</b> TD	NT	AA Tongth	Score	Probability
ORF Name	NTID	AAID	Length	Length		
29973182_c2_246	322	5544	857	2574	1042	3.4e-105
Protein name				Locus	Name	Acc#
(p)ppGpp synthetas	se			gp:BSU	36377	U86377
Description						
Bacillus subtilis adeninephosphoribos		_			cds.	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
3131910_c2_268	323	5545	492	1479	832	6.0e-83
Protein name				Locus	Name	Acc#
				sp:YCG	_ECOLI	P76007
Description					0	
PUTATIVE NA(+)/H(-	+) EXCHANGE	R YCGO				, .
			NT	AA		
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
ORF Name 3165800c3302		<u>AAID</u>	Length		Score	Probability  [5.5e-09
			Length	Length		
3165800c3302			Length	Length	134 Name	5.5e-09
3165800 <u>c3</u> 302 Protein name			Length	Length  Locus	134 Name	5.5e-09 <u>Acc#</u>
3165800c3302	324	5546	Length	Length Locus Locus gp:AB01	134 Name 15879	5.5e-09 <u>Acc#</u>
Protein name ORF2 Description	324	5546	Length  103	Length Locus gp:AB01	134 S Name 15879	5.5e-09 Acc# AB015879
Protein name ORF2 Description	324	5546	Length	Length Locus Locus gp:AB01	134 Name 15879	5.5e-09 <u>Acc#</u>
Protein name ORF2 Description Porphyromonas ging	givalis dnal	5546 K operon c	Length  103  genes, co  NT  Length	Length Locus Locus pp:AB01	134 S Name 15879	5.5e-09 Acc# AB015879
Protein name ORF2 Description Porphyromonas gine ORF Name	givalis dnal	K operon c	Length  103  Jenes, co	Length Locus gp:AB01 complete	134 S Name 15879 Cds.	Acc# AB015879  Probability
Protein name ORF2 Description Porphyromonas ging ORF Name 32660011_f3_150	givalis dnal	K operon c	Length  103  Jenes, co	Length Locus gp:AB01 complete	134 S Name 15879 Score 276 S Name	5.5e-09  Acc#  AB015879  Probability  5.8e-24
Protein name ORF2 Description Porphyromonas ging ORF Name 32660011_f3_150	givalis dnal	K operon c	Length  103  Jenes, co	Length Locus  In price of the control of the contro	134 S Name 15879 Score 276 S Name	5.5e-09  Acc#  AB015879  Probability  5.8e-24  Acc#

Part of Andrew Street S	ji m
Part of Andrew Street S	₹# 163
The second secon	4, 5
1948 S.	
1948 S.	190 see
THE RESIDENCE BETWEEN THE BEST STATES	
THE RESIDENCE OF THE REAL PROPERTY.	
THE RESIDENCE OF THE REAL PROPERTY.	# 55
10.00 1 10.00 1 10.00 1 10.00 1 10.00 1 10.00 1 10.00 1 10.00 1 10.00 1 10.00 1 10.00 1 10.00 1 10.00 1 10.00 1	1 14
	4 14 4 14 4 15
	OR DAY
45 155 4 155 2 14	72
F 1942	122
M	F 1942
in we.	

ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability	
33240828_f2_62	326 55	48	885	2658			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT				<del> </del>			]
			NT	AA			<del>-</del> -
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
3.3.78.83.8.7£27.8	327 55	49	310	933	101	0.024	
Protein name				Locus	Name	Acc#	
conserved hypothet	ical protein	10		pir:G72	385	G72385	
Description				I			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
33992211 <u></u> c1181	328 55	50	132	399	200	5.6e-16	
Protein name				Locus	Name	Acc#	
diacylglycerol kin	ase			gp:BSU2	9177	U29177	
Description							
Bacillus subtilis	PhoH (phoH) g	ene, par	rtial co	s, diacy	lglycero	olkinase (dgk)	7
gene, complete cds,		_			_	-	
			177	7.7			_
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
34178385 <u></u> £3 <u></u> 149	329 55	51	511 1	.536	2540	6.1e-264	_
Protein name				Locus	Name	Acc#	
				sp:ATPE	BACFR	P13356	
Description					· · · · · · · · · · · · · · · · · · ·	-	
ATP SYNTHASE BETA	CHAIN.			·			1
							j

	NTID AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name			1809	452	8.9e-52
34181561_£1_19	330 5552				
Protein name				Name	Acc#
115K outer memb protein	orane protein precurso	or:SusC	pir:JC6	5027	JC6027
Description			_		
		NT	AA		- 1 1 1 1 1 1 1
ORF Name	NTID AAID	<u>Length</u>	Length	Score	Probability
3.56.250.00 <u></u> £2 <u>92</u>	331 5553	530	1593		
Protein name			Locus	s Name	Acc#
Description					
NO-HIT					
		NT	AA	Score	Probability
ORF Name	NTID AAID	Length	<u>Length</u>	BCOLE	11030311107
3937751_f2_101		864	2595	404	1.3e-36
Protein name			Locu	s Name	Acc#
DNA helicase ho	omolog		gp:AF1	08138	AF108138
Description					
Homo sapiens D	NA helicase homolog (	PIF1) mRNA	, partia	l cds.	
		NT	<u>AA</u>	Score	Probability
ORF Name	NTID AAID	Length	Length	50020	
43.95252 <u></u> c1 <u></u> 209	333 5555	528	1587	494	1.7e-62
Protein name			Locu	s Name	Acc#
Beta-N-Acetylg	lucosaminidase		gp:AB0	15350	AB015350
Description				1	199
Streptomyces to complete cds.	hermoviolaceus nagB g	ene forBet	a-N-Acet	yIgIucos	aminidase,

			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length	20010	
4454637_f1_5	334	5556	315	948	1058	6.8e-107
Protein name				Locus	Name	Acc#
dTDP-glucose 4-6-	_	:protein		pir:S75	550	S75550
slr0809:protein sl	10809					
Description						
			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>	<u> </u>	11000011111
4501875 <u>f2</u> 51	335	5557	124	375		
Protein name				Locus	Name	Acc#
Description					!	
NO-HIT			-			
110-1111						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
4553376c1199						
# 7. 7. 7. 4. (1) 10 4 1. 4. 4. 4	336	5558	459	L380	291	2.5e-25
Protein name	336	5558	459		291 Name	2.5e-25 <u>Acc#</u>
	336	5558	459 [1		Name	
	336	5558	459	Locus	Name	Acc#
Protein name				Locus	Name	Acc#
Protein name  Description			minutes.	Locus gp:ECOU	Name	Acc# L10328
Protein name  Description				Locus	Name	Acc#
Protein name  Description  E. coli; the regi	on from 81.	5 to 84.5	minutes.  NT Length	Locus gp:ECOU	Name	Acc# L10328
Protein name  Description  E. coli; the region  ORF Name	on from 81.	5 to 84.5 <u>AAID</u>	minutes.  NT Length	Locus  gp:ECOU  AA  Length	Name W82 Score	Acc# L10328  Probability
Protein name  Description  E. coli; the region  ORF Name  4881660_c2_266	on from 81.	5 to 84.5 <u>AAID</u>	minutes.  NT Length	Locus  gp:ECOU  AA  Length	Score  [1809] Name	Acc# L10328  Probability  1.8e-186
Protein name  Description  E. coli; the region  ORF Name  4881660_c2_266	on from 81.	5 to 84.5 <u>AAID</u>	minutes.  NT Length	Locus  AA  Length  Locus	Score  [1809] Name	Acc# L10328  Probability  1.8e-186  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
5276557 f2 98	338	5560	169 5	10	197	1.2e-15
Protein name			<u> </u>	Locus	s Name	Acc#
ATP synthase F0, s	ubunit b			pir:H7	2231	H72231
Description						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
5276586 <u></u> ±2 <u></u> 99	339	5561	189	570	215	1.4e-17
Protein name				Locu	s Name	Acc#
F1F0-ATPase subuni	t delta			gp:AF0	98522	AF098522
Description						
Lactobacillus acid partial cds; and F1	ophilus ur F0-ATPase	acil phos operon, c	phoribosy complete s	yltransf sequence	erase (u <sub>]</sub>	op)gene,
	· · · · · · · · · · · · · · · · · · ·	-	NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>		6.4e-06
556557 <u></u> ±3 <u></u> 134	340	5562	478	1437	136	
Protein name				Locu	s Name	Acc#
				sp:YF0	7_METJA	Q58902
Description						-
HYPOTHETICAL PROTE	EIN MJ1507					
			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length		
5890667 <u>_</u> c1_176	341	5563	95	288	213	2.4e-17
Protein name					ıs Name	Acc#
RNA-binding prote	in			gp:ANA	RBPD2	D49425
Description					e.	
	is rbpD ger					

ORF Name	NTID AA	<u>NT</u> ID Length	AA Length Score	Probability	
6046881_f3_140	342 5564		921 655	3.4e-64	
Protein name			Locus Name	Acc#	
3-methyl-2-oxobut	anoate		gp:CGPAN	X96580	
Description					
C.glutamicum pani	3, panC & xylB ge	nes.			
ORF Name	NTID AA	<u>NT</u> ID <u>Length</u>	AA Length Score	Probability	
6836013_c2_265	343 5565	443	1332 281	1.1e-21	
Protein name			Locus Name	Acc#	
			sp:NTRY_AZOCA	Q04850	
Description					
NITROGEN REGULAT	ION PROTEIN NTRY,				
NITROGEN REGULAT		<u>NT</u> AID Length	AA Length Score	Probability	0
	NTID AA	NT Length	- Score	Probability	
ORF Name	NTID AA	NT Length	Length Score	Probability  Acc#	
ORF Name 7.0.7.2.0.3.7	NTID AA	NT Length	Length Score 723	, * x	
ORF Name 7.072037_c3_322	NTID AA	NT Length	Length Score 723	, * x	
ORF Name 7.072037c3322  Protein name Description	NTID AA	NT Length	Length Score  Locus Name	Acc#	
ORF Name 7.072037c3322  Protein name Description	NTID AA	NT Length	Length Score  Locus Name	, * x	
ORF Name 7.0.7.2.0.3.7	NTID AA 344   5566	NT Length  240  NT Length  NT Length	Length Score  Locus Name  AA Score	Acc# Probability	
ORF Name 7.0.7.2.0.3.7	NTID AA 344   5566	NT Length  240  NT Length  NT Length	Length Score  Locus Name  AA Length Score	Acc#	
ORF Name 7.0.7.20.3.7	NTID AA 344   5566	NT Length  240  NT Length  NT Length	Length  723  Locus Name  AA Length  Score  Score	Acc# Probability	





			$\overline{ ext{NT}}$	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length	<del></del>	
9766325_f3_148	346	5568	422	.269	1070	3.6e-108
Protein name				Locus	s Name	Acc#
				sp:PUR	r_syny3	Q55336
Description					ľ	
2) (FORMATE-DEP	ENDENT GAR TR	RANSFORMYL	ASE)			
				7. 70		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
1057762_f2_175	347	5569	214	545	528	9.8e-51
Protein name		-		Locu	s Name	Acc#
thio-specific a	ntioxidant (t	sa) perox	idase	pir:E7	2036	E72036
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name		<u>AAID</u>	Length		Score	Probability
ORF Name  1173263c2422  Protein name			Length	Length	Score s Name	Probability  Acc#
1173263 <u></u> c2 <u>422</u>			Length	Length	1	
1173263 <u></u> c2 <u></u> 422			Length	Length	1	
1173263 <u></u> c2 <u>422</u>			Length	Length	1	
Protein name  Description			Length	Length	s Name	Acc#
Protein name  Description			Length	Length 234 Locu	1	
Protein name  Description  NO-HIT	348 NTID	5570	Length  NT Length	Length  234  Locu	s Name	Acc#
Protein name  Description  NO-HIT  ORF Name	348 NTID	5570 AAID	Length  77  NT  Length	Length  Locu  AA  Length  900  Locu	Score 79 IS Name	Acc#  Probability  0.021  Acc#
1173263c2422  Protein name  Description  NO-HIT  ORF Name  11834382c2479.	NTID 349	5570 AAID	Length  77  NT  Length	Length  Locu  AA  Length	Score 79 IS Name	Acc# Probability 0.021
Protein name  Description  NO-HIT  ORF Name  11834382_c2_479.  Protein name	NTID 349	5570 AAID	Length  77  NT  Length	Length  Locu  AA  Length  900  Locu	Score 79 IS Name	Acc#  Probability  0.021  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
118906_c1_393	350	5572	165	498		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
12204402 <u></u> c2 <u></u> 483	351	5573	67	204	72	0.020
Protein name		- 2	<del></del>	Locus	Name	Acc#
pE66L				gp:ASU1	8466	U18466
Description						
African swine fev	er virus, co	mplete g	enome.			* 1
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
12677202 <u></u> £3 <u></u> 307	352	5574	237	714	472	8.5e-45
Protein name				Locus	Name	Acc#
hypothetical prot	ein			gp:AHAA	MYG	X58627
Description						1
A.haloplanktis am 1,4-alpha-D-glucan			Lase		3	* *
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	<u>50010</u>	TIODADILICY
13103938 <u></u> c2 <u></u> 466	353	5575	159	180	227	5.6e-22
Protein name				Locus	Name	Acc#
single stranded D	NA-binding p	rotein		gp:SSU6	4095	U64095
Description						
Shewanella sp. PT cds.	99 single st	randed D	NA-bindir	ng protei	n (ssb)	gene, complete

ORF Name	NTID AAID	NT AA Score Length Length	Probability
1367792_f3_258	354 5576	202 609 298	2.3e-26
Protein name		Locus Name	Acc#
		sp:YB69_HAEIN	P44118
Description			
HYPOTHETICAL PRO	OTEIN HI1169		·
		NT AA G	Drobobility
ORF Name	NTID AAID	Length Length Score	Probability
1369082_c1_320	355 5577	419 1260 714	3.8e-75
Protein name		Locus Name	Acc#
autoaggregation-	-mediating protein	gp:AF091502	AF091502
Description			
Lactobacillus re	euteri autoaggregation	n-mediating protein (aggH	gene,
complete cds.			
-		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
13.78.78.27 <u></u> c2 <u></u> 46.8		156 471	
Protein name		Locus Name	Acc#
Description			
NO-HIT		. *	
		NT AA Grana	Probability
ORF Name	NTID AAID	Length Length Score	Probability
13876713 <u>_</u> £2 <u>_</u> 124	357 5579	103 312	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
LAO. 117.	0.00		

ORF Name	NTID AAID	NT AA Length Length Score	Probability
13914808_£2_181	358 5580	103 312	
Protein name		Locus Name	Acc#
Description			×
NO-HIT			
ORF Name	NTID AAID	NT AA Score	Probability
14304550 <u>c2</u> 490	. 359 5581	380 1143	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT <u>AA</u> Length Length <u>Score</u>	Probability
1441937c1386		322   969	** * * *
Protein name		Locus Name	Acc#
Description		-	
NO-HIT			
	2270	NT <u>AA</u> Score Length Length	Probability
ORF Name 14494037c3524	NTID AAID 5583		1.5e-32
Protein name		Locus Name	Acc#
		sp:CCSA_CYACA	P31564
Description			
CYTOCHROME C BIOG	ENESIS PROTEIN CCSA	*	

ORF Name	NTID AAID	NT AA Score Probability Length Length
	362 5584	1249 3750 259 7.8e-18
Protein name		Locus Name Acc#
		gp:SCYDL057W
Description		
S.cerevisiae chrom	nosome IV reading fr	came ORF YDL057w.
ORF Name	NTID AAID	NT AA Score Probability Length Length
14665882_c2_461	363 5585	153 462
Protein name		Locus Name Acc#
Description		
NO-HIT		
ORF Name	NTID AAID	NT AA Score Probability
14875191 <u></u> c1388	364 5586	365 1098 116 0.00061
Protein name		Locus Name Acc#
hypothetical prot	ein	gp:YEN132945 AJ132945
Description	-	
Yersinia enteroco	litica WA 314 right	arm of the high-pathogenicityisland.
ORF Name	NTID AAID	NT AA Score Probability Length Length
156325 <u>f</u> 2_176	365 5587	116 351 211 3.8e-17
Protein name		Locus Name Acc#
ss-DNA binding pr	otein 12RNP2 precur	sor gp:SYO12RNP2 D17359
Description		
Synechococcus 630	)1 gene for ss-DNA k	pinding protein 12RNP2, completecds.

ORF Name	NTID AAID	<u>NT</u> Length	AA Score	Probability
15660937_c1_345	366 5588	432 1	299 617	3.6e-60
Protein name			Locus Name	Acc#
hypothetical prote	in		pir:T33724	T33724
Description				
		NTITI	7.7	
ORF Name	NTID AAID	<u>NT</u> Length	<u>AA</u> <u>Length</u> <u>Score</u>	Probability
162785_c3_565	. 367 5589	322 9	156	7.4e-09
Protein name			Locus Name	Acc#
Mag44			gp:DEPMAG44	D17682
Description				
Dermatophagoides fa	arinae mRNA for Mag	44. parti	al cds.	<u> </u>
			AA	
ORF Name	NTID AAID	NT Length	AA Length Score	Probability
	NTID AAID	<u>NT</u> Length	AA Score	Probability
ORF Name	NTID AAID	<u>NT</u> Length	AA Length Score	Probability  Acc#
ORF Name 16447875G3560	NTID AAID	<u>NT</u> Length	AA Length Score	
ORF Name  16.4478.753560  Protein name	NTID AAID	<u>NT</u> Length	AA Length Score	
ORF Name  16447875_c3_560  Protein name  Description	NTID AAID	NT Length	AA Score Length  Locus Name	
ORF Name  16447875_c3_560  Protein name  Description	NTID AAID	<u>NT</u> Length	AA Length Score	
ORF Name  16447875c3560  Protein name  Description  NO-HIT	NTID AAID 368 5590	NT Length  63 II	AA Score	Acc#
ORF Name  16447875c3560  Protein name  Description  NO-HIT  ORF Name	NTID AAID  368 5590  NTID AAID	NT Length  63 II	AA Score Length  Locus Name  AA Score  Locus Name	Acc#
ORF Name  16447875c3560  Protein name  Description  NO-HIT  ORF Name  16525318c2471	NTID AAID  368 5590  NTID AAID	NT Length  63 II	AA Score Length  Locus Name  AA Score  Locus Name  AA Score  Length	Acc# Probability
ORF Name  16.44.78.75	NTID AAID  368 5590  NTID AAID	NT Length  63 II	AA Score Length  Locus Name  AA Score  Locus Name  AA Score  Length	Acc# Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
165882_c1_371	370	5592	1018	3057	179	7.8e-10
Protein name				Locus	Name	Acc#
				sp:PRIM	_CLOAB	P33655
Description						
DNA PRIMASE,						
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
16900087_c2_420	371	5593	279	840	414	1.2e-38
Protein name				Locus	s Name	Acc#
hypothetical prot	tein yycJ		-8	pir:A70	0090	A70090
Description						
			NT	ΔΔ		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 17010890_c1_369		<u>AAID</u> 5594	Length	_	Score	Probability
			Length	Length	Score	Probability  Acc#
17010890 <u></u> c1 <u></u> 369			Length	Length		
17010890c1369 Protein name			Length	Length		
Protein name  Description			Length  113	Length  342  Locus		Acc#
Protein name  Description			Length	Length		
Protein name  Description  NO-HIT	NTID	5594	Length  113  NT Length	Length  342  Locus	s Name	Acc#
Protein name  Description  NO-HIT  ORF Name	NTID	5594 AAID	Length  113  NT Length	Length  Locus  AA  Length	s Name	Acc#
Protein name  Description  NO-HIT  ORF Name  180302_c3_559	NTID	5594 AAID	Length  113  NT Length	Length  Locus  AA  Length	s Name	Acc# Probability
Protein name  Description  NO-HIT  ORF Name  180302_c3_559  Protein name	NTID	5594 AAID	Length  113  NT Length	Length  Locus  AA  Length	s Name	Acc# Probability  Acc#

ORF Name	NTID AAID	NT AA Score Probability Length Length
187930_c3_509	374   5596	359 1080
Protein name		Locus Name Acc#
Description		
NO-HIT	-	
ORF Name	NTID AAID	NT AA Score Probability Length Length
19035 <u></u> £1 <u></u> 47	375 5597	674 2025 1563 2.1e-160
Protein name		Locus Name Acc#
branching enzyme		gp:AB026630 AB026630
Description		
Emericella nidular	ns gene for branchi	ng enzyme, complete cds.
		NT AA Score Probability
ORF Name	NTID AAID	Length Length Score Probability
197191 <u>    f3</u> _291	376 5598	381 1146 129 3.1e-05
Protein name	* 	Locus Name Acc#
		sp:PORP_PSEAE P05695
Description		
PORIN P PRECURSOR	OUTER MEMBRANE PR	ROTEIN D1)
PORIN P PRECURSOR	. (OUTER MEMBRANE PR	NT AA
ORF Name	OUTER MEMBRANE PR	
	NTID AAID	NT AA Score Probability
ORF Name	NTID AAID	NT AA Score Probability Length Length
ORF Name	NTID AAID	NT AA Score Probability Length Length  171 516
ORF Name 19729591_cl_380  Protein name	NTID AAID	NT AA Score Probability Length Length  171 516

ODE Namo	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name		5600		582	96	0.021	_
19734688_c2_423	378	5600			L		
<u>Protein name</u>					Name	Acc#	
two-component sens	sor histidin	e kinase	homolog	pir:F69	747	F69747	
Description				2			
			NT	AA		p	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
19740893 <u></u> c3 <u></u> 542	379	5601	408	L227	1169	1.2e-118	
Protein name				Locus	Name	Acc#	
				sp:FSR	_ECOLI	P52067	
Description							
FOSMIDOMYCIN RESI	STANCE PROT	EIN	)			~ "	7
			NT	AA	~	Duchabilita	
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
ORF Name 1988388_t1_44			Length		Score	Probability	
4444		AAID	Length	Length	Score	Probability  Acc#	
1988388 <u>_</u> £1 <u></u> 44		AAID	Length	Length			
1988388_f1_44 Protein name		AAID	Length	Length	- 8		
Protein name  Description		AAID	Length	Length	s Name	Acc#	<u>-</u>
Protein name  Description		AAID	Length	Length  195  Locus	- 8		
1988388_f1_44 Protein name Description	NTID	<u>AAID</u> 5602	Length  64  NT  Length	Length  195  Locus	s Name	Acc#	
1988388_f1_44 Protein name Description NO-HIT ORF Name	NTID	AAID 5602 AAID	Length  NT Length	Length  Locus  AA  Length	S Name	Acc# Probability	
1988388_f1_44 Protein name Description NO-HIT ORF Name 20314007_c2_407	NTID	AAID 5602 AAID	Length  NT Length	Length  Locus  AA  Length  2484  Locus	Score	Acc# Probability 4.8e-184	
1988388_f1_44 Protein name Description NO-HIT ORF Name 20314007_c2_407	NTID	AAID 5602 AAID	Length  NT Length	Length  Locus  AA  Length  2484  Locus	Score 1786 S Name	Acc#  Probability  4.8e-184  Acc#	

			NT	AA	Score	Probability	
ORF Name	$\underline{\mathtt{NTID}}$	AAID	<u>Length</u>	Length			
20344086_f2_157	382	5604	65	.98			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NTC	7.7			_
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
20350260 <u></u> c2 <u></u> 405	383	5605	70	213			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT				-			
			איד	AA		- 1 1 2 2 2	_
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 204437t3277		<u>AAID</u> 5606	Length		Score	Probability	
			Length	Length	Score S Name	Probability  Acc#	
20.443.7 <u></u> £3 <u></u> 27.7			Length	Length			
204437 <u>f3</u> 277			Length	Length			
Protein name  Description			Length	Length 696  Locu		Acc#	
Protein name  Description			Length	Length			
Protein name  Description  NO-HIT	384 NTID	5606	Length  231  NT  Length	Length 696  Locus	s Name	Acc#	
Protein name  Description  NO-HIT  ORF Name	384 NTID	5606	Length  231  NT  Length	Length  Locus  AA  Length	s Name	Acc#	
Protein name  Description  NO-HIT  ORF Name  20501551_c1_325	384 NTID	5606	Length  231  NT  Length	Length  Locus  AA  Length	s Name	Acc# Probability	

thraft.	22d	
i,	Percent Section	
1000	1 20	
	my	
7	Hear.	
W	DIA.	
22		
E. Land	die die	
1	121 122	
7	825	
100	azi	
133	Head of the same o	
133		
The stands of		
The stands of		
The stands of		
一部 明明 田 明明 田		
The stands of the stands of the stands		
The stands of the stands of the stands		

ORF Name	NTID AAID	NT AA Score Length Length	Probability
20980313_c1_392	386 5608	69 210 47	0.034
Protein name		Locus Name	Acc#
		sp:YOR5_TTV1	P19280
Description			
HYPOTHETICAL 9.5	KD PROTEIN		
		77	
ORF Name	NTID AAID	NT AA Score Length Length	Probability
2126506_c1_314	387 5609	170 513 93	0.00016
Protein name		Locus Name	Acc#
transcription reg	Julator phage-related	homolog pir:C69774	C69774
ydcn			
Description			
	<u> </u>		
ORF Name	NTID AAID	NT AA Score Length Length	Probability
21485027 <u></u> c2 <u></u> 481	388 5610	192 579	**
Protein name		Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Score Length Length	Probability
21490925c1365	389   5611	110 333	
Protein name		Locus Name	Acc#
		· ·	
Description			
NO-HIT			

			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length		
21664055_c3_585	390	5612	81 2	246	69	0.042
Protein name				Locus	s Name	Acc#
ATP synthase gamma	chain			gp:AB02	27877	AB027877
Description						
Schizosaccharomyces	s pombe ge	ne for AT	P synthas	se gamma	chain,pa	artial cds,
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	*	
21677180_c3_566	391	5613	106	321		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT		-	*			
NO-HIT		-	NT	AA	Score	Probability
NO-HIT ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
		<u>AAID</u> 5614	Length		Score	Probability  0.031
ORF Name			Length	Length		
ORF Name 21688925f140.			Length	Length	56 s Name	0.031
ORF Name  21688925f140  Protein name			Length	Length Locus	56 s Name	0.031 <u>Acc#</u>
ORF Name  21688925f140  Protein name  estrogen receptor			Length	Length  Locu  Locu  pir:S2	56 s Name	0.031 <u>Acc#</u> S26595
ORF Name  21688925f140  Protein name  estrogen receptor			Length	Length Locus	56 s Name	0.031 <u>Acc#</u>
ORF Name  21688925f140.  Protein name  estrogen receptor  Description	NTID	5614	Length  66  NT Length	Length Locu Locu Dir:S2	56 S Name 6595	0.031 <u>Acc#</u> S26595
ORF Name  21688925f140  Protein name  estrogen receptor  Description  ORF Name	NTID	5614 <u>AAID</u>	Length  NT Length	Length Locus Locus  pir:S2  AA Length	Score	0.031  Acc#  S26595  Probability
ORF Name  21688925f140  Protein name  estrogen receptor  Description  ORF Name  21729812c2408	NTID 393	<u>AAID</u>	Length  NT Length	Length Locus Locus  pir:S2  AA Length	Score 236 s Name	0.031 Acc# S26595  Probability 1.0e-17

A SE
40 14 sa
1,5
22 421 E25
41.00
Brate and
4 10
The state of the s
žį.
122
21 197
f
227
31 agr
IJ

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
21756268_c2_419	394	5616	81	246		
Protein name	<u>                                  </u>	<u> </u>	<u> </u>	Locus	s Name	Acc#
Description						
NO-HIT	· · · · · · · · · · · · · · · · · · ·					
			2.77	7.7		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2206.7162t153	395	5617	506	521	388	7.9e-41
Protein name				Locus	Name	Acc#
				sp:GLNA	A_BACCE	P19064
Description				<u> </u>		
	'ASE. (GLUTA	MATE AMM	ONTA LTG	ASE)		*
Description GLUTAMINE SYNTHET	ASE, (GLUTA	MATE AMM	ONIA LIGA	ASE)		
	'ASE, (GLUTA NTID	MATEAMM AAID	ONIA LIGA <u>NT</u> Length	ASE)  AA  Length	Score	Probability
GLUTAMINE SYNTHET	NTID		<u>NT</u> Length	AA Length	Score	Probability
GLUTAMINE SYNTHET  ORF Name  22087762c3496	NTID	AAID	<u>NT</u> Length	AA Length		4
ORF Name  22087762_c3_496  Protein name	NTID	AAID	<u>NT</u> Length	AA Length	Score S Name	Probability  Acc#
GLUTAMINE SYNTHET  ORF Name  22087762c3496	NTID	AAID	<u>NT</u> Length	AA Length		4
ORF Name  22087762_c3_496  Protein name	NTID	AAID	<u>NT</u> Length	AA Length		4
ORF Name  2208.7.62_c3_496  Protein name  Description	NTID	AAID	NT Length	AA Length 570 Locus		4
ORF Name  2208.7.62_c3_496  Protein name  Description	NTID	AAID	<u>NT</u> Length	AA Length		4
GLUTAMINE SYNTHET  ORF Name  22087762_c3_496  Protein name  Description  NO-HIT	NTID 396 NTID	<u>AAID</u> 5618	NT Length 189 F	AA Length  Tocus  AA	3 Name	Acc#
GLUTAMINE SYNTHET  ORF Name  2208.7.7.62	NTID 396 NTID	AAID 5618 AAID	NT Length 189 F	AA Length  Locus  AA Length	3 Name	Acc#
GLUTAMINE SYNTHET  ORF Name  2208.7.7.62	NTID 396 NTID	AAID 5618 AAID	NT Length 189 F	AA Length  Locus  AA Length	Score	Acc# Probability
ORF Name  22087762_c3_496  Protein name  Description  NO-HIT  ORF Name  22132811_f2_180  Protein name	NTID 396 NTID	AAID 5618 AAID	NT Length 189 F	AA Length  Locus  AA Length	Score	Acc# Probability

•	NTID AAID	NT AA Score	Probability
ORF Name			3.3e-75
22459802_f2_159	398 5620		
Protein name		Locus Name	Acc#
p-aminobenzoate syn	nthase component I	homolog pir:F64187	F64187
Description			÷
ORF Name	NTID AAID	NT AA Score	Probability
22462807 <u>c2476</u>		178   537   93	0.042
		Locus Name	Acc#
<u>Protein name</u>		sp:TGN3_RAT	P19814
		sp:1GN3_KA1	
Description			
TRANS-GOLGI NETWOR	K INTEGRAL MEMBRANE	PROTEIN TGN38 PRECURSOR	
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
22656553 <u></u> ±3 <u></u> 296	400 5622	275 828	
Protein name		Locus Name	Acc#
Description			
NO-HIT		* .	
		NT AA Grama	
ORF Name	NTID AAID	Length Length Score	Probability
23438887 <u>_c1</u> _389	401 5623	126 381 87	0.013
Protein name		Locus Name	Acc#
unknown	*	gp:AF074396	AF074396
Description			
Desulfotomaculum	thermocisternum	(myss) cono no	rtial cdg:
karda ferredoxin (fo	dx). dissimilatory :	ansferase (murA) gene, pa sulfite reductase subunit	A
(dsrA), dissimilato:	ry sulfite reductas	e subunit B (dsrB), and d	srD
genes, complete cds	; and unknown gene.		

			$\underline{\mathbf{NT}}$	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>	<u> </u>	120000022207
23472178_f3_271	402	5624	470	L413	1205	1.8e-122
Protein name				Locus	s Name	Acc#
Xylose Isomerase				gp:RFL	132472	AJ132472
Description						
Ruminococcus flav	efaciens xyl	an utili	zation op	peron.		
			2777	7.7		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
23597202_c3_513	403	5625	198	597	48	0.039
Protein name				Locu	s Name	Acc#
hypothetical prot	ein F21D9.3			pir:T2	1205	T21205
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
23.63.2132£16.4	404	5626	510	1533	614	7.6e-60
Protein name		<u> </u>		Locu	s Name	Acc#
xylulose kinase	* -			gp:AF0	01974	AF001974
Description	š -	ו				
Thermoanaerobacte	r ethanolicu	s putati	ve TrkG	gene, pa	rtial cds	s, andputative
TrkA, xylose isome	rase (xylA)	and xylu	lose kin	ase (xyl	B)genes,	complete cds.
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
23.63.47.08c13.68	405	5627	84	255	69	0.042
Protein name				Locu	s Name	Acc#
				sp:YC1	3_METJA	Q58610
Description						
HYPOTHETICAL PROT	EIN MJ1213					-

ORF Name	NTID AAID		ore Probability
23651680_±2_191	406 5628	78 237	
Protein name		Locus Na	me <u>Acc#</u>
Description			
NO-HIT			
ORF Name	NTID AAII	NT AA Scott Length	ore Probability
236.908.75 <u></u> c1 <u></u> .318	407 5629	385 1158 931	1.9e-93
Protein name		Locus Na	me Acc#
		sp:TGT_BAC	SU 032053
Description			*
	) (GUANINE INSERTIC	ONT TIMESYMEN	
TRANSGLYCOSYLASE	(COMMINE INSERTE	ON ENZIME)	
TRANSGLYCOSYLASE	(GOANINE INSERTIO		
ORF Name	NTID AAII	NT AA Sc	ore Probability
	NTID AAII	NT AA Sc	ore Probability  0.013
ORF Name	NTID AAII	NT AA Sc Length Length	0.013
ORF Name 23.710.926f.2184	NTID AAII	NT AA Sc Length Length Sc 355 1068 82	0.013 me <u>Acc#</u>
ORF Name  23.710.926f2184  Protein name	NTID AAII	NT AA SC Length Length Sc 355 1068 82 Locus Na	0.013 me <u>Acc#</u>
ORF Name  23.710926f2184  Protein name  M protein precur	NTID AAII	NT AA SC Length Length SC Locus Na pir:S61081	0.013 me <u>Acc#</u>
ORF Name  23.710926f2184  Protein name  M protein precur	NTID AAII	NT AA SC Length Length SC Locus Na pir:S61081	0.013 me <u>Acc#</u>
ORF Name  23.710.926f2184  Protein name  M protein precur  Description	NTID AAII 408 5630 sor NTID AAI	NT AA SC Length Length B2  Locus Na pir:S61081	0.013 me <u>Acc#</u>
ORF Name  23.710.926f218.4  Protein name  M protein precur  Description  ORF Name	NTID AAII 408 5630 sor NTID AAI	NT AA SC Length Length 82  Locus Na pir:S61081  NT AA Length Length Sc Length Length Sc	me Acc#  ore Probability
ORF Name  23.710.926f2184  Protein name  M protein precur  Description  ORF Name  23.86.56.51c1327	NTID AAII 408 5630 sor NTID AAI	NT AA SC Length Length Sc Locus Na pir:S61081  NT AA SC Locus Na D Length Length Sc Length Length Sc Length Length Sc Length Sc Length Sc Length Length Length Sc Length Length Length Sc Length Length Sc Length Length Sc Length Length Sc Length Length Length Sc Length Sc Length Length Sc Length	me Acc#  ore Probability

1	1007
- 57	Graff Han
100,000	Fin
	Sitt
331	
Signalia Signalia	5FL
100	ij
H 23	200 200 200 200 200 200 200 200 200 200
ħ,	
淵	
記が記	27. 27.
	st:
ass	200
Ľ	
## ##	
Straigs	
7	

		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
23922135_c3_495	410 5632	133 402 73	0.016
Protein name		Locus Name	Acc#
MesF		gp:AF143443	AF143443
Description			
Leuconostoc meser mesentericin B105	nteroides plasmid pHY (mesB), MesH (mesH),	30 MesG (mesG) gene, par and MesF (mesF)genes, c	omplete cds.
ORF Name	NTID AAID	<u>NT AA</u> Score Length Length	Probability
24027213_c2_460  Protein name	411 5633	206 621 Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Score	Probability
ORF Name 24241261f177		Length Length Score	
ORF Name		Length Length Score	Probability  Acc#
ORF Name 24241261f177		Length Length Score	
ORF Name  24241261f177  Protein name		Length Length Score	
ORF Name  24241261f177  Protein name  Description		Length Length Score	
ORF Name  24241261_f1_77  Protein name  Description  NO-HIT  ORF Name  24265886_c2_469	412   5634	Length Length    229   690     Locus Name	Acc# Probability
ORF Name  24241261_f1_77.  Protein name  Description  NO-HIT  ORF Name	412   5634	Length Length    229   690     Locus Name	Acc#
ORF Name  24241261_f1_77  Protein name  Description  NO-HIT  ORF Name  24265886_c2_469	412   5634	Length Length    229   690     Locus Name	Acc# Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24305437_c3_556	414	5636	80	243			
Protein name			<u> </u>	Locus	Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24322127 <u></u> c3 <u></u> 56.7	415	5637	202	609			
Protein name				Locus	Name	Acc#	
Description				•			
NO-HIT							
			NT	AA	<del></del>		
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability	
24344641£3233	416	5638	120	363			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT					= i.c	*	
			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
24397877 <u>c1</u> 374	417	5639	64	195	71	0.0075	
Protein name				Locus	Name	Acc#	
				sp:RAFR	ECOLI	P21867	
				DP VIIII			
Description					ī		

		NT AA	Score	Probability
ORF Name	NTID AAID	Length Length		
24406557_f2_158	418 5640	252 759	151	4.2e-08
Protein name		<u>Loc</u>	us Name	Acc#
protein antigen Lms	STI1	gp:LM	U73845	U73845
Description				
Leishmania major p	rotein antigen LmS	TII mRNA, partia	I cds.	
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Lengtl	Score	Probability
	419 5641	122 369	80	0.0029
Protein name		Loc	us Name	Acc#
putative repressor	protein	gp:BA	1242593	AJ242593
Description				
Bacteriophage All8	complete genome.			
ORF Name	NTID AAID	<u>NT AA</u> Length Lengt	h Score	Probability
24484682 <u></u> c2 <u></u> 492	420 5642	359 1077	]	
Protein name	,	Loc	cus Name	Acc#
	· .	Loc	cus Name	Acc#
Description	,	Loc	cus Name	Acc#
		Loc	cus Name	Acc#
Description	NTID AAID	<u>NT</u> <u>AA</u>	Score	Acc# Probability
Description NO-HIT ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Lengt	h Score	
Description NO-HIT ORF Name 24503282_c3_538		NT AA Length Lengt	h Score	Probability  0.017
Description NO-HIT  ORF Name 24503282_c3_538	421 5643	NT AA Length Lengt  83 252 Loc	h Score 78 US Name	Probability  0.017  Acc#
Description NO-HIT ORF Name 24503282_c3_538	421 5643	NT AA Length Lengt  83 252 Loc	h Score	Probability  0.017

ODE Nama	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name					C3.	4.5e-62
24633387_c1_354	422	5644	331 9	96	635	4.56-62
Protein name				Locus	s Name	Acc#
hypothetical protein	n T27E13.6	5		pir:T0	0580	T00580
Description						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
24640915c2480	423	5645	157 4	74		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT		, , , , , , , , , , , , , , , , , , , ,				-
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
24644636 <u></u> c1 <u></u> 3.72	424	5646	111 3	36		
Protein name				Locu	s Name	Acc#
Description						
NO-HIT						
			) TITL	7. 7.		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24645337 <u></u> ±3 <u></u> 295	425	5647	439	.320	1245	1.0e-126
Protein name				Locu	s Name	Acc#
putative UDP-glucos	e dehydro	genase		gp:AF1	59428	AF159428
Description			= "			
Burkholderia pseudo ADP-heptose synthase (gmhD) genes, comple	e (waaE),	tative UD and putat	P-glucose iveADP-gl	e dehydr Lycero-m	ogenase annohepto	(udg),putative ose epimerase

			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length	-	
24648412_f1_23	426	5648	166 5	01		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24648562 <u></u> c3 <u></u> 555	427	5649	82 2	249		
Protein name		,		Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
24710926 <u></u> £3 <u></u> 299	. 428	5650	174	525	121	1.8e-07
Protein name				Locus	Name	Acc#
thiol:disulfide in	cerchange	protein h	omolog	pir:E69	9891	E69891
yneN						
Description						
		· · · · · · · · · · · · · · · · · · ·	NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>	BCOIC	11000011101
24782635 <u></u> ±1 <u></u> 93	. 429	5651	187	564	544	2.0e-52
Protein name				Locus	s Name	Acc#
dTDP-6-deoxy-D-glu		nimerase		gp:AF0	48749	AF048749
	cose-3,5 e	Pimerabe		] gp		
Description	cose-3,5 e	Pimerase	-	95		

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24798568_£2_219	430	5652	206	621.			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							7
							J 
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24804681_c1_370	431	5653	74	225			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT	· · · · · · · · · · · · · · · · · · ·						7
							J 
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 24804807t2160		<u>AAID</u> 5654	Length	_	Score	Probability	<b>.</b>
			Length	Length	Score Name	Probability  Acc#	X.
24804807t2_160			Length	Length			<b>.</b>
24804807f2160 Protein name			Length	Length			,
Protein name  Description			Length	Length 587 Locus			× ×
Protein name  Description			Length	Length			× .
Protein name  Description  NO-HIT	432 NTID	5654	Length  228  NT  Length	Length 587 Locus	Name	Acc#	, , , , , , , , , , , , , , , , , , ,
Protein name  Description  NO-HIT  ORF Name	432 NTID	AAID	Length  228  NT  Length	Length Locus  AA Length	Name	Acc#	, , , , , , , , , , , , , , , , , , ,
24804807t2160  Protein name  Description  NO-HIT  ORF Name  24820925t180	432 NTID	AAID	Length  228  NT  Length	Length Locus  AA Length	Name Score	Acc# Probability	× .

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24882942 c3_544	434	5656	357	1074	666	2.3e-65
			L L	T.OCUS	Name	Acc#
Protein name						032223
				sp:YVA	A_BACSU	032223
Description						
HYPOTHETICAL OXI	DOREDUCTASE	IN FHUD-O	PUBD INT	ERGENIC F	REGION	
			NT	AA	G	Drobabilitu
ORF Name	NTID	AAID	Length	Length	Score	Probability
2537802 c2_404	435	5657	181	546	52	0.044
Protein name			[ L	Locus	s Name	Acc#
1	o Foin			gp:AF0		AF021739
T				l ab. Hro.	21733	
envelope glycopr						
Description						
Description HIV-1 isolate sl	ng clone 45	from the	Netherla	nds, enve	elopegly	coprotein
Description	ng clone 45	from the	Netherla	nds, enve	elopegly	coprotein
Description HIV-1 isolate sl	ng clone 45	from the	NT	AA		
Description HIV-1 isolate sl	ng clone 45	from the			elopegly Score	rcoprotein Probability
Description HIV-1 isolate sl (env) gene, parti	ng clone 45 al cds. NTID		NT Length	AA		
Description HIV-1 isolate sl (env) gene, parti ORF Name	ng clone 45 al cds. NTID	AAID	NT Length	AA Length		
Description HIV-1 isolate sl (env) gene, parti ORF Name 25428312_f3_285	ng clone 45 al cds. NTID	AAID	NT Length	AA Length	Score	Probability
Description HIV-1 isolate sl (env) gene, parti ORF Name	ng clone 45 al cds. NTID	AAID	NT Length	AA Length	Score	Probability
Description HIV-1 isolate sl (env) gene, parti ORF Name 25428312_f3_285	ng clone 45 al cds. NTID	AAID	NT Length	AA Length	Score	Probability
Description  HIV-1 isolate sl (env) gene, parti  ORF Name  25428312_f3_285  Protein name  Description	ng clone 45 al cds. NTID	AAID	NT Length	AA Length	Score s Name	Probability Acc#
Description  HIV-1 isolate sl (env) gene, parti  ORF Name  25428312_f3_285  Protein name  Description	ng clone 45 al cds. NTID	AAID	NT Length	AA Length 525 Locu	Score	Probability
Description  HIV-1 isolate sI (env) gene, parti  ORF Name  25428312_f3_285  Protein name  Description  NO-HIT  ORF Name	ng clone 45 al cds.  NTID  436	<u>AAID</u>   5658	NT Length 174	AA Length 525 Locu AA Length	Score s Name	Probability Acc#
Description HIV-1 isolate sI (env) gene, parti  ORF Name  25428312_f3_285  Protein name  Description  NO-HIT  ORF Name  25556532_c2_402	ng clone 45 al cds.  NTID  436	AAID 5658	NT Length  174  NT Length	AA Length  AA Locu  AA Length	Score s Name	Probability Acc#
Description  HIV-1 isolate sI (env) gene, parti  ORF Name  25428312_f3_285  Protein name  Description  NO-HIT  ORF Name	ng clone 45 al cds.  NTID  436	AAID 5658	NT Length  174  NT Length	AA Length  AA Locu  AA Length	Score Score	Probability  Acc#  Probability
Description HIV-1 isolate sI (env) gene, parti  ORF Name  25428312_f3_285  Protein name  Description  NO-HIT  ORF Name  25556532_c2_402	ng clone 45 al cds.  NTID  436	AAID 5658	NT Length  174  NT Length	AA Length  AA Locu  AA Length	Score Score	Probability  Acc#  Probability

			NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length			
25572212_c1_315	438	5660	122 3	69	69	0.042	
Protein name				Locus	Name	Acc#	
hypothetical prot	ein yopO			pir:Tl	2849		
Description							
			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
25664086 <u></u> c1 <u></u> 390	439	5661	303	12			
Protein name				Locus	s Name	Acc#	
Description							
NO-HIT	-						1
							<u> </u>
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
2595058_f2_127	440	5662	70 2	213			
Protein name			- %	Locu	s Name	Acc#	
Description							
NO-HIT							]
			NT	AA			<u> </u>
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
26433216_c3_499	441	5663	81 2	246			5
Protein name			1	Locu	s Name	Acc#	
Description							
NO-HIT	*		<u> </u>		**		1
			NT	AA			
ORF Name	NTID	AAID	Length	<u>AA</u> Length	Score	Probability	
26517 <u>c1</u> 346	442	5664	69	210			
Protein name				Locu	s Name	Acc#	
Description							
NO-HIT	*	-					

NT AA C D L L L L L L L L L L L L L L L L L L
ORF Name NTID AAID Length Length Score Probability
26601510_c2_448 443 5665 71 216 101 1.7e-05
Protein name Locus Name Acc#
hypothetical protein MJ1608 pir:G64500 G64500
Description
NIII 22
ORF Name NTID AAID Length Length Score Probability
26601577 <u>c1</u> 359
Protein name Locus Name Acc#
conserved hypothetical protein aq_1386 pir:F70420 F70420
Description
NT AA .
ORF Name NTID AAID Length Length Score Probability  [266.923.42
266.923.42c2446
Protein name Locus Name Acc#
succinate CoA ligase (ADP-forming), beta pir: H70439 H70439
chain
Description
OPE Name NULL AND LONG Probability
ORF Name NIID AAID Length Length
273261c2_477
Protein name Locus Name Acc#
<u>Description</u>
NO-HIT
NT AA Coore Deckehilita
ORF Name NTID AAID Length Length Score Probability
27.703.05c3580
Protein name Locus Name Acc#
Description
NO-HIT

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
2822161_c1_395  Protein name	448	5670	733	2202 Locus	: Name	Acc#	
Description				and the same of th	~		1
NO-HIT							j
ORF Name	NTID	<u>AAID</u>	NT Length	AA Length	Score	Probability	
Protein name	449	5671	73	Locus	: Name	Acc#	
Description							_
NO-HIT					*		
	*	- 0	NTITI	7) 7)			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
2923187 <u></u> c2 <u></u> 424	450	5672	149	450	138	2.1e-09	
Protein name				Locus	Name	Acc#	
				sp:YG02	HAEIN	P44270	
Description						1	
HYPOTHETICAL PROTEI	N HI1602						
		<u> </u>	NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
29339125 <u>_</u> f2 <u>_</u> 122	451	5673	79	240			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT				-			]

ORF Name NTID AAID Length Length Probabili	ty,
29412901_f1_79	
Protein name Locus Name Acc	<u> </u>
sp:LSPA_STACA Q598	35
Description	
PEPTIDASE) (SIGNAL PEPTIDASE II) (SPASE II)	1 2
	ity
29470081_c1_376	
Protein name Locus Name Acc	<u>#</u>
hypothetical protein PH0283 pir:D71453 D714	53
Description	
NT AA G	
ORF Name NTID AAID Length Length Probabil:	ıty
3.028.9001c3525454 5676 468 1407 1084 1.2e-10	)9
Protein name Locus Name Acc	<u>#</u>
cytochrome c peroxidase gp:AF200362 AF20	0362
Description	
Haemophilus ducreyi oxaloacetate decarboxylase gamma chain (oadG)gene, partial cds; oxaloacetate decarboxylase alpha chain (oadA),oxaloacetate decarboxylase beta chain (oadB), and alkylphosphonateuptake protein (phn genes, complete cds; ccp gene, completesequence; cytochrome c peroxidase gene, complete cds; and unknowngene.	a)
NT AA Ggorg Brobabil	
ORF Name NTID AAID Length Length Score Probabil:	ity
3.058.8453 <u>f11</u>	
Protein name Locus Name Acc	<u>#</u>
Description	
NO-HIT	

	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name		5678		789	118	0.00032
31672502_f1_24	456	5676	202			
Protein name					Name	Acc#
type I restriction protein H91_orf543:h	enzyme hsd	M:hypothe	etical	pir:S7	3820	S73820
H91_orf543				i		
Description						
			NT	AA		Probability
ORF Name	NTID	AAID	Length	Length	Score	Probability
32035967_c3_507	457	5679	121	366	161	7.6e-12
Protein name				Locu	s Name	Acc#
hypothetical prote	in			gp:SSU	18930	Y18930
Description						*
Sulfolobus solfata	ricus 281	kb genomi	c DNA fr	agment,	strain P	2.
20.220		J		,		
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
	NTID		NT Length	AA		
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
ORF Name 3.3.3.64.7.62	<u>NTID</u>	<u>AAID</u> 5680	NT Length	AA Length	Score 890 s Name	Probability 4.3e-89
ORF Name 33364762_c1_356	<u>NTID</u>	<u>AAID</u> 5680	NT Length	AA Length 903 Locu	Score 890 s Name	Probability  4.3e-89  Acc#
ORF Name  3.3.3.64762_c1_3.56  Protein name  succinateCoA light	<u>NTID</u>	<u>AAID</u> 5680	NT Length	AA Length 903 Locu	Score 890 s Name	Probability  4.3e-89  Acc#
ORF Name  3.3.3.64762_c1_3.56  Protein name  succinateCoA lighter chain	<u>NTID</u>	<u>AAID</u> 5680	NT Length	AA Length 903 Locu	Score 890 s Name 9719	Probability  4.3e-89  Acc#  F69719
ORF Name  3.3.3.64762_c1_3.56  Protein name  succinateCoA lighter chain	<u>NTID</u>	<u>AAID</u> 5680	NT Length 300	AA Length  903  Locu  pir:F6	Score 890 s Name	Probability  4.3e-89  Acc#
ORF Name  3.3.3.647.62c13.56  Protein name  succinateCoA light chain  Description	NTID  458  ase (ADP-fo	AAID 5680 orming),	NT Length  300  alpha  NT Length	AA Length  903  Locu  pir:F6	Score 890 s Name 9719	Probability  4.3e-89  Acc#  F69719
ORF Name  3.3.3.64762c13.56  Protein name  succinateCoA lighter  chain  Description  ORF Name	NTID  458  ase (ADP-fo	AAID  STRING  AAID	NT Length 300 alpha  NT Length	AA Length  903  Locu  pir:F6  AA Length	Score  890  s Name  9719  Score	Probability  4.3e-89  Acc#  F69719  Probability
ORF Name  3.3.3.647.62c13.56  Protein name  succinateCoA lighter  chain  Description  ORF Name  3.3.85.955c13.60	NTID  458  ase (ADP-fo	AAID  STRING  AAID	NT Length 300 alpha  NT Length	AA Length  903  Locu  pir:F6  AA Length	Score  890  s Name  9719  Score  148  s Name	Probability  Acc# F69719  Probability  8.5e-16

	NIM TO	70 77 TT	NT	AA Longth	Score	Probability	
ORF Name	NTID	AAID	Length	Length			
34017140_c3_498	460	5682	64	195			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
34073250f154	461	5683	258	777	234	1.4e-19	
Protein name				Locus	Name	Acc#	
				sp:YT29	_MYCTU	P71564	
Description							
PUTATIVE OXIDOREDU	CTASE RV09	45,					
			NT	AA			
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
ORF Name 3.417.646.22418		<u>AAID</u>	Length	<del></del>	<u>Score</u>	Probability 7.5e-140	
			Length	Length			
3.417.64.62 <u></u> c2 <u></u> 418			Length	Length	1369 Name	7.5e-140	
3.417.64.62 <u></u> c2 <u></u> 418			Length	Length 1419 Locus	1369 Name	7.5e-140	
34176462c2418 Protein name			Length	Length 1419 Locus	1369 Name	7.5e-140	
Protein name  Description			Length 472	Length  1419  Locus  sp:UXAC	1369 Name	7.5e-140 <u>Acc#</u>	
Protein name  Description			Length	Length 1419 Locus	1369 Name	7.5e-140	
Protein name  Description  ISOMERASE)	MTID	5684	Length  A72  NT Length	Length Locus Sp:UXAC	Name ECOLI	7.5e-140 <u>Acc#</u>	
Protein name  Description  ISOMERASE)  ORF Name	MTID	AAID	Length  A72  NT Length	Length Locus Sp:UXAC	Name ECOLI	7.5e-140 <u>Acc#</u>	
13.41.7.64.6.2	MTID	AAID	Length  A72  NT Length	Length Locus Sp:UXAC	Name ECOLI Score	7.5e-140 Acc# Probability	
Protein name  Description  ISOMERASE)  ORF Name  34277280_c1_378	MTID	AAID	Length  A72  NT Length	Length Locus Sp:UXAC	Name ECOLI Score	7.5e-140 Acc# Probability	

ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
34406502_c2_403  Protein name	464	5686	130	Locus	3 Name	Acc#
Description				······································		
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
34407787 <u> </u> £1 <u> </u> 55	465	5687	198	597	132	7.8e-07
Protein name	_			Locus	Name	Acc#
				sp:Y374	_METJA	Q57819
Description						-16
HYPOTHETICAL PROT	EIN MJ0374					
						t e
1111 01112110111 11101						
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
	NTID	<u>AAID</u> 5688	Length		Score	Probability  3.9e-127
ORF Name	NTID		Length	Length		
ORF Name 35634587 <u></u> t1 <u></u> 74	NTID		Length	Length	1249 Name	3.9e-127
ORF Name 35634587 <u></u> t1 <u></u> 74	NTID		Length	Length  Locus	1249 Name	3.9e-127 <u>Acc#</u>
ORF Name 35634587_t1_74 Protein name Description	<u>NTID</u> 466	5688	Length	Length 1599  Locus	1249 Name BACSU	3.9e-127 <u>Acc#</u>
ORF Name 35634587 <u>174</u> Protein name	<u>NTID</u> 466	5688	Length	Length 1599  Locus	1249 Name BACSU	3.9e-127 <u>Acc#</u>
ORF Name 35634587_t1_74 Protein name Description	<u>NTID</u> 466	5688	Length	Length 1599  Locus	1249 Name BACSU	3.9e-127 <u>Acc#</u>
ORF Name 35634587f174  Protein name  Description  HYPOTHETICAL 60.2	NTID 466  KD PROTEIN NTID	IN CSPB-C	Length  532  LPP INT  NT  Length	Length Locus Sp:YHCX ERGENIC R	1249 Name BACSU EGION	3.9e-127 Acc# P54608
ORF Name 35634587f174  Protein name  Description  HYPOTHETICAL 60.2	NTID 466  KD PROTEIN NTID	IN CSPB-C	Length  532  LPP INT  NT  Length	Length Locus Sp:YHCX ERGENIC R AA Length	Name BACSU EGION Score	3.9e-127  Acc#  P54608  Probability
ORF Name 35634587f174  Protein name  Description  HYPOTHETICAL 60.2  ORF Name 3633.7562c3553	NTID  466  KD PROTEIN  NTID  467	IN CSPB-C	Length  532  LPP INT  NT  Length	Length Locus Sp:YHCX ERGENIC R AA Length	Name BACSU EGION Score 91 Name	3.9e-127  Acc# P54608  Probability  0.00020
ORF Name 35634587tl74  Protein name  Description  HYPOTHETICAL 60.2  ORF Name 3633.7562c3553	NTID  466  KD PROTEIN  NTID  467	IN CSPB-C	Length  532  LPP INT  NT  Length	Length Locus sp:YHCX ERGENIC R AA Length Locus	Name BACSU EGION Score 91 Name	3.9e-127  Acc# P54608  Probability  0.00020  Acc#

Hand.	Hank But Bank
¥	225
The state of the s	Total S
	***
4	731
1	1. m
100	Merce.
Hard.	100 m
100	A
500	
Specie.	* 8 San
20	222
-	magn:
Sept.	· 是 化" 是 是
	222
37.	*****
15449	H., 12
the Hard House	##. ##

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
36500003_c1_328	468	5690	430	293	849	9.5e-85	]
Protein name				Locus	Name	Acc#	
macrolide-efflux c	leterminant			gp:SPU8	3667	U83667	
Description							
Streptococcus pneucds.	imoniae mac	rolide-ef	flux dete	erminant	(mefE) o	gene, complete	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	-
3938838_c1_373	469	5691	131	396			
Protein name				Locus	Name	Acc#	
Description					11		
NO-HIT							
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability	
		<u>AAID</u> 5692	Length		Score	Probability	
ORF Name		G	Length	Length	Score Score	Probability  Acc#	
ORF Name 4021882_c3_563		G	Length	Length	*		
ORF Name 4021882c3563		G	Length	Length	*		
ORF Name  4021882_c3_563  Protein name  Description		G	Length  431	Length  Locus	3 Name		
ORF Name  4021882_c3_563  Protein name  Description		G	Length  431  NT Length	Length  Locus  AA  Length	*	Acc#	
ORF Name  4021882c3563  Protein name  Description  NO-HIT	470 NTID	5692	Length  431  NT Length	Length Locus  AA Length	Score	Acc# Probability	
ORF Name  4021882c3563  Protein name  Description  NO-HIT  ORF Name	470 NTID	5692 AAID	Length  431  NT Length	Length Locus  AA Length	3 Name	Acc#	
ORF Name  4021882c3563  Protein name  Description  NO-HIT  ORF Name  4062628c3577	470 NTID	5692 AAID	Length  431  NT Length	Length Locus  AA Length	Score	Acc# Probability	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4065760_f1_63	472	5694	255	768	239	4.1e-20
Protein name				Locus	Name	Acc#
hypothetical protei	n			pir:S75	926	S75926
Description						
			NTT.	7) 7)		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4072687 <u>_</u> f1 <u>_</u> 81	473	5695	773	2322	121	1.5e-05
Protein name				Locus	s Name	Acc#
outer membrane prot	ein			gp:NGU	31959	U81959
Description						
Neisseria gonorrhoe	ae outer n	nembrane p	protein	(omp85) s	gene, cor	mpletecds.
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
4.1.5.7.5.3.6 <u></u> c3 <u></u> 5.7.6	474	5696	88	267	77	0.018
Protein name			-	Locu	s Name	Acc#
hypothetical protei	n ZC47.1			pir:T2	7592	T27592
Description				<u> </u>		100
		0				
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4318885 <u>f2174</u>	475	5697	509	1530	1371	4.6e-140
Protein name	-				s Name	Acc#
xylose transporter				gp:AB0	09593	AB009593
Description						
Tetragenococcus had partial and complete		osC, rbsB	, xylR,	xylA, xy	lB and x	ylEgenes,

	NTITL'T TO	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name	NTID			591	92	0.028	7
4345012_f2_166	476	5698	190		L		
Protein name				Locus	Name	Acc#	
				sp:CRP_	ECOLI	P03020	
Description							
PROTEIN)							
ORF Name	NTID	AAID	<u>NT</u> <u>Length</u>	<u>AA</u> Length	Score	Probability	
4553288_f1_45	477	5699	67	204			
Protein name			-	Locus	Name	Acc#	
Description							
NO-HIT							i
F							1
			אויוי	ΔΔ			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
		<u>AAID</u> 5700	Length		Score	Probability	
ORF Name			Length	Length	Score S Name	Probability  Acc#	6
ORF Name 46.9.5.28.5			Length	Length			8
ORF Name 46.95285_c2_434  Protein name			Length	Length			=
ORF Name 46.95285_c2_434  Protein name  Description			Length	Length 960  Locus		Acc#	=
ORF Name 46.95285_c2_434  Protein name  Description			Length	Length			=
ORF Name  46.95285_c2_434  Protein name  Description  NO-HIT		5700	Length 319  NT Length	Length  Jocus	s Name	Acc#	
ORF Name  46.9.5.2.8.5		5700 <u>AAID</u>	Length 319  NT Length	Length  Locus  AA  Length	s Name	Acc#	
ORF Name  46.95285c2434  Protein name  Description  NO-HIT  ORF Name  47.73400c3557		5700 <u>AAID</u>	Length 319  NT Length	Length  Locus  AA  Length	s Name Score	Acc# Probability	

ORF Name	NTID AAII	NT AA Score Probability D Length Length
4791400_f2_134  Protein name	480 5702	71 216  Locus Name Acc#
Description		
NO-HIT		
ORF Name	NTID AAII	NT AA Score Probability  Length Length
4877135 <u></u> £3 <u></u> 279	481 5703	1149 3450 991 1.7e-211
Protein name		Locus Name Acc#
isoleucinetRNA l synthetase:isoleucy		
Description		
ORF Name	NTID AAII	NT AA Score Probability D Length Length
48.83.5.92 <u></u> c3 <u></u> 5.7.4	482 5704	181 546
Protein name		Locus Name Acc#
Description		
NO-HIT	7	
ODE Name	NTID AAI	NT AA Score Probability  ID Length Length
ORF Name [49.783.02c3504		424   1275   659   1.3e-64
Protein name		Locus Name Acc#
probable phosphose	erine phosphatase	pir:T36772 T36772
Description		
	-	
ORF Name	NTID AAI	
564077 <u></u> c2 <u>44</u> 7	484 5706	78 237
Protein name		Locus Name Acc#
Description		
NO-HIT		

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
6022037_c1_337	485	5707	353 1	.062	764	9.6e-76
Protein name				Locus	s Name	Acc#
				IHY:qa	- ECOLI	
Description						
HYPOTHETICAL 39.2	KD PROTEIN	IN RHSB-	PIT INTER	RGENIC RE	EGION	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
6347188_c3_588	486	5708	1643	1932	161	1.3e-08
Protein name				Locus	s Name	Acc#
				gp:AB0	08550	AB008550
Description						
Pseudomonas aerug	inosa phage	phi CTX.	complete	genome	sequence	2.
	F	Para Carry	00p00.	2 901100		
		P111 0111,				
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
	NTID		<u>NT</u> Length	AA	L.	
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name  6.8.3.4.8.0.7t.17.8	NTID 487	AAID	<u>NT</u> Length	AA Length	Score 127 s Name	Probability  3.1e-08
ORF Name  6834807t178  Protein name	NTID 487	AAID	<u>NT</u> Length	AA Length 393	Score 127 s Name	Probability  3.1e-08  Acc#
ORF Name  6.8.348.0.7t178  Protein name  probable dnak sup	NTID 487	AAID	NT Length	AA Length  393  Locus	Score 127 s Name	Probability  3.1e-08  Acc#
ORF Name  6.8.348.0.7t178  Protein name  probable dnak sup	NTID 487	AAID	<u>NT</u> Length	AA Length 393	Score 127 s Name	Probability  3.1e-08  Acc#
ORF Name  6.8.348.0.7t178  Protein name  probable dnaK sup  Description	NTID 487  pressor  NTID	<u>AAID</u> 5709	NT Length  130  NT Length	AA Length  B93  Locus  pir:D7	Score 127 S Name	Probability  3.1e-08  Acc#  D71366
ORF Name  5.8.348.0.7t178  Protein name  probable dnak sup  Description  ORF Name	NTID 487  pressor  NTID	<u>AAID</u> 5709	NT Length  130	AA Length  Bocus  Locus  pir:D7  AA Length	Score  127  S Name 1366  Score	Probability  3.1e-08  Acc#  D71366  Probability
ORF Name  6.8.348.0.7£178	NTID 487  pressor  NTID 488	<u>AAID</u> 5709	NT Length  130	AA Length  Bocus  Locus  pir:D7  AA Length	Score  127  S Name  1366  Score  283  S Name	Probability  3.1e-08  Acc#  D71366  Probability  9.0e-25

鬱	1475
Eroji.	100
115	
١.	The state of
, and	2.5
Sector 5	M
ŝ,	1 8
	11/2
21	1100
ą,	25
F	100
41	
-	diam'r.
5	: 3
15.00 FF	100
ž,	
**	
5	1165°
-93	1107
ă,	1199
	1192,
**	1165
H	
H	
	145 145 145
H	
H	
****	100 100 100 100 100 100 100 100 100 100
****	
# W.#	
****	A STATE OF S
****	
****	A STATE OF S
****	
****	
****	4.2
新 いま ま さっぱ ま まっき	4.2
新 いま ま さっぱ ま まっき	4.2
****	

		NT	AA Score	Probability
ORF Name	NTID AA		Length —	
829436_c3_535	489 5711	343	600	2.3e-58
Protein name			Locus Name	Acc#
protein kinase homo	log Thi		gp:AF070520	AF070520
Description				
Sinorhizobium melil	oti protein ki	nase homolog	Thi (thi) andEx	oP-like
protein genes, compl	ete cds; and u	nknown genes	•	-
		NT	AA Score	Probability
ORF Name	NTID AA	ID <u>Length</u>	Length	
84637_c1_366	490 5712	68	207	
Protein name			Locus Name	Acc#
Description				
NO-HIT				
NO IIII				
ORF Name	NTID AA	NT Length	<u>AA</u> Length Score	Probability
9884427_£3_287	[491   5713	66	201	* .
Protein name	J		Locus Name	Acc#
FIOCEIT Hame				*
Description				
NO-HIT				
		NT	AA Score	Probability
ORF Name	NTID A	<u>NT</u> AID Length	AA Length Score	Probability
ORF Name 9.95.7.827_c1_3.26		AID Length	Score	Probability  3.3e-14
		AID Length	Length Score	-
9.95.7.8.27c13.2.6		AID Length	Length Score	3.3e-14
9.95.7.8.27c13.2.6		AID Length	Length Score  Length 110  Locus Name	3.3e-14 <u>Acc#</u>

ą.	227	
	ane Fare	
ķ	arê	
45.00	Ŋ	
	atgi;	
¥!	ali.	
	SII.	
100	1 "	
Brass Real		
21	115	
ŧ,		
100	1111	
į,	455	
帯		
Į.	114. 114.	
ŀ,		
	1100 1100	
77	44	
i di	HEL.	
4		
	17425	
F)	11423	
N.	Haral St.	
þ		
No.	1100	
'n	11 die	

ORF Name	NTID AAID	NT AA Score Length Length	Probability
10320312_f2_49  Protein name	493 5715	328 987  Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Score Length Length	Probability
10681577 <u></u> c3 <u></u> 251	494 5716	103 312 114	7.3e-07
Protein name		Locus Name	Acc#
hypothetical prot	cein APE1165	pir:H72586	H72586
Description			· · · · · · · · · · · · · · · · · · ·
ORF Name	NTID AAID	$rac{ ext{NT}}{ ext{Length}}$ $rac{ ext{AA}}{ ext{Length}}$ Score	Probability
10828567 <u></u> £3 <u></u> 89	495 5717	204 615 148	1.8e-10
Protein name		Locus Name	Acc#
conserved hypothe	etical protein	pir:C72361	C72361
Description			
ORF Name	NTID AAID	NT AA Score	Probability
13865887c2187	496 5718	64 195	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	1.3e-32
1414187 <u></u> £3 <u></u> 80	497 5719	316 951 357	
Protein name		Locus Name	Acc# AB012956
		gp:AB012956	AB012330
Description		synthesis, strain MO45,c	omplete cds.
Vibrio cholerae	genes for U-antigen	Synchesis, Schain nois,	

ORF Name	NTID AAID	NT AA Score	Probability
14485841 f3 95	498 5720	208 627 659	1.3e-64
Protein name		Locus Name	Acc#
rubrerythrin	*	gp:AF202316	AF202316
Description			
Moorella thermoa	acetica rubrerythrin	gene, complete cds.	
		NT AA	
ORF Name	NTID AAID	- Score	Probability
14491537_f1_22	499 5721	682 2049 101	0.012
Protein name		Locus Name	Acc#
comEA protein-re	:lated protein	pir:F72301	F72301
Description			
		NT AA	
ORF Name	NTID AAID	- Score	Probability
14569387 <u></u> £3 <u></u> 81	500 5722	158 477	
Protein name	× *	Locus Name	Acc#
Description			
NO-HIT			
		NT AA Grana	
ORF Name	NTID AAID	— Score	Probability
15105001 <u></u> c2 <u></u> 156	501 5723	119 360 87	0.013
Protein name		Locus Name	Acc#
hypothetical pro	otein M70.1	pir:T33032	Т33032
Description			(1)
		NT AA Grava	
ORF Name	NTID AAID	- Score	Probability
15.751503 <u></u> 228	502 5724	219 660	
Protein name		Locus Name	Acc#
Description			
NO-HIT			e

See 15	in the second
1	
-	Time.
	1000
瓣	(C)
	HALL THE
il.	100
12 cas 34	100 P
ŧ,	222. 221.
#	
il.	HOLE HALE
	120
357	H4S
Hand	
	per l
est.	: P.
Street,	140. 140. 140.
1	HE R

ODE Name	NTID AAID	NT AA Score Probability Length Length
ORF Name 158136_f3_76	503   5725	308 927 399 4.6e-37
Protein name		Locus Name Acc#
conserved hypotheti	cal protein	pir:G72409 G72409
Description		
		NT AA Goro Probability
ORF Name	NTID AAID	NT AA Score Probability Length Length
16.056463 <u></u> 236	504 5726	75 228
Protein name		Locus Name Acc#
Description		
NO-HIT		
		NT AA Score Probability
ORF Name	NTID AAID	Length Length
194025_f1_30	505 5727	145 438
Protein name		Locus Name Acc#
Description		
NO-HIT		
		NT AA Score Probability
ORF Name	NTID AAID	Length Length 3.5e-16
20506875 <u></u> £3 <u></u> 79	5728	
Protein name		Locus Name Acc#
*		sp:Y516_BORBU 051468
Description	<del>ሉኩእናለ - አለርመከነፕ</del> የሃኛ <b>ም</b> ስ አእ <u>ኛ</u> ላ ውር	PACE BROSIA
HYPOTHETICAL TRNA/	RKNA METHYLIKANSFE	IADE DECOTO

Single Mary	100
107	1907
i,	E57.
	122
Market St.	Sec.
100	
	2965
190	250
音が始	Service Servic
80	227
No.	1 3
Ĩ	1
St.	
14	rije Tija
1.	報
1	2
4	
15-11	325
ij,	CH2
	227
183	1,000
23	1212.
No. 19	227
	242
134	3113
37	Manie Manie
F.	100 m
ici	100
Sand	and.

ORF Name	NTID A	<u>NT</u> AID Length	AA Score	Probability
20509632_£1_31	507 5729	9 448	1347 829	1.2e-82
Protein name			Locus Name	Acc#
dihydrolipoamide dehydrogenase,:2-0	xoglutarate dehv	vdrogenase	pir:140794	140794
complex chain E3:a	cetoin dehydroge	enase complex	لع	
Description				
	NTID A	<u>NT</u> AID Length	AA Score	Probability
ORF Name	N11D A		1252	
20995143_c2_182		0 0	Locus Name	Acc#
Protein name			Hoods Hame	
Description				
NO-HIT				
ORF Name	NTID A	<u>NT</u> AID Length	AA n Length	Probability
ORF Name 2220010_c1_147		AID Length	Score	Probability
-		AID Length	n <u>Length</u> Score	Probability  Acc#
2220010 <u></u> c1 <u></u> 147		AID Length	n Length Score	.:
2220010c1147		AID Length	n Length Score	.:
Protein name  Description		AID Length	Locus Name	Acc#
Protein name  Description	509 573	AID Length	Locus Name  AA Score	Acc#
Protein name  Description  NO-HIT	509 573	AID Length  1 82  NT AID Length	Locus Name  AA Score	Acc#
Protein name  Description  NO-HIT  ORF Name	509 573	AID Length  1 82  NT AID Length	Locus Name  AA Score	Acc# Probability
Protein name  Description  NO-HIT  ORF Name  223.82311f13	NTID A	AID Length  1 82  NT AID Length	Locus Name  Locus Name  AA Length  Score	Acc# Probability  3.4e-25
Protein name  Description  NO-HIT  ORF Name  22382311f13  Protein name	NTID A	AID Length  1 82  NT AID Length	Locus Name  AA Score Length  Locus Name  Locus Name  Locus Name  Locus Name	Acc#  Probability  3.4e-25  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
22679637_±1_11	511	5733	199	500	366	1.4e-33	
Protein name				Locus	s Name	Acc#	
conserved hypothet	ical prote	in ysnA		pir:C6	9986	C69986	
Description	_						
			NT	AA		01-14-19-1914	
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
23516942 <u></u> f1 <u></u> 15	. 512	5734	493	L482	127	8.3e-05	
Protein name				Locus	s Name	Acc#	
outer membrane pro	tein tolc ]	precursor	(tolC)	pir:H7	1733	H71733	
Description							
			N7/77	70.70			
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
ORF Name 23594000_f1_17		<u>AAID</u>	Length	*	Score	Probability	
107			Length	Length	Score S Name	Probability  Acc#	
23594000 <u>f1</u> 17			Length	Length	e *2	*	
23594000f117 Protein name			Length	Length	e *2	*	
23594000_f1_17Protein name  Description			Length  192	Length  579  Locus	e *2	*	
23594000_f1_17Protein name  Description			Length	Length	e *2	*	-
Protein name  Description  NO-HIT	. [513	5735	Length  192  NT Length	Length  Locus	s Name	Acc#	
Protein name  Description  NO-HIT  ORF Name	NTID		Length  192  NT Length	Length Locus  AA Length	Score	Acc# Probability	
23594000f117	NTID		Length  192  NT Length	Length  Locus  AA  Length  Locus  Locus	Score	Acc# Probability 3.0e-129	
23594000f117	NTID		Length  192  NT Length	Length  Locus  AA  Length  Locus  Locus	Score  [1269] S Name	Acc#  Probability  3.0e-129  Acc#	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
23986267_f1_19  Protein name	515	5737	148	Locus	s Name	Acc#
Description						0 - 11
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24.116.56.7t253	516	5738	457	1374	194	2.5e-12
Protein name				Locus	Name	Acc#
chromosomal hemolys:	in D			gp:AF08	31284	AF081284
Description						\$
Escherichia coli st cds; and Hp1 (hp1),	rain CFT0 Hp2 (hp2)	73 chromos , Hp3 (hp:	somal her 3), and 1	molysin D Hp4 (hp4)	genes, (	gene,partial complete cds.
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
242253.02 <u>f1</u> 16	517	5739	607	1824		
Protein name				Locus	Name	Acc#
Description						
NO-HIT	···	· · · · · · · · · · · · · · · · · · ·		·		
-			NT	AA		
ORF Name	NTID	AAID	Length	<u>AA</u> Length	Score	Probability
24260952 <u></u> c3 <u></u> 221	518	5740	84	255		
Protein name				Togue	Name	
				посия	Name	Acc#
Description	-	÷		Locus	Name	ACC#

		NT AA .
ORF Name	NTID AAID	- Score Propability
24409662_c2_166	519 5741	104 315 107 6.8e-11
Protein name		Locus Name Acc#
iron-uptake fact	or	gp:AF051690 AF051690
Description		
		factor (piuC), eptor (piuA), and iron-uptake factor
ORF Name	NTID AAID	NT AA Score Probability
24415875_f2_55	520 5742	538 1617 521 7.2e-68
Protein name		Locus Name Acc#
arylsulfatase		gp:PAATSAGN Z48540
Description		*
Pseudomonas aeru	ginosa atsR, atsB, a	atsC & atsA genes.
	- · · · · · · · · · · · · · · · · · · ·	aced a depit genes.
ORF Name	NTID AAID	NT AA Score Probability
ORF Name 24495337_c2_190	NTID AAID	NT AA Score Probability
	NTID AAID	NT AA Score Probability
	NTID AAID	NT AA Score Probability Length Length 962 1.0e-96
	NTID AAID	NT AA Score Probability  Length Length 962 1.0e-96  Locus Name Acc#
24495337c2190 Protein name	NTID AAID	NT AA Score Probability  Length Length 962 1.0e-96  Locus Name Acc#
Protein name  Description	NTID AAID	NT AA Score Probability  338 1017 962 1.0e-96  Locus Name Acc#  sp:NADA_SYNY3 P74578
Protein name  Description	NTID AAID	NT AA Score Probability    NT Length Length   Score Probability     338   1017   962   1.0e-96     Locus Name   Acc#     Sp:NADA_SYNY3   P74578
Protein name  Description  QUINOLINATE SYNT	NTID AAID 521 5743 HETASE A NTID AAID	NT AA Score Probability    NT Length Length   Score Probability     338   1017   962   1.0e-96     Locus Name   Acc#     Sp:NADA_SYNY3   P74578
Protein name  Description  QUINOLINATE SYNT	NTID AAID 521 5743 HETASE A NTID AAID	NT AA Score Probability    NT Length Length   Score Probability     338   1017   962   1.0e-96     Locus Name   Acc#     Sp:NADA_SYNY3   P74578     NT AA Length Length   Score Probability
Protein name  Description  QUINOLINATE SYNT  ORF Name	NTID AAID 521 5743 HETASE A NTID AAID	NT Length         AA Length         Score         Probability           338         1017         962         1.0e-96           Locus Name         Acc#           sp:NADA_SYNY3         P74578           NT Length         Score         Probability           Length         Length         5.7e-39
Protein name  Description  QUINOLINATE SYNT  ORF Name	NTID AAID 521 5743 HETASE A NTID AAID	NT Length         AA Length         Score         Probability           338         1017         962         1.0e-96           Locus Name         Acc#           sp:NADA_SYNY3         P74578           NT Length         AA Length         Score         Probability           474         1425         417         5.7e-39           Locus Name         Acc#

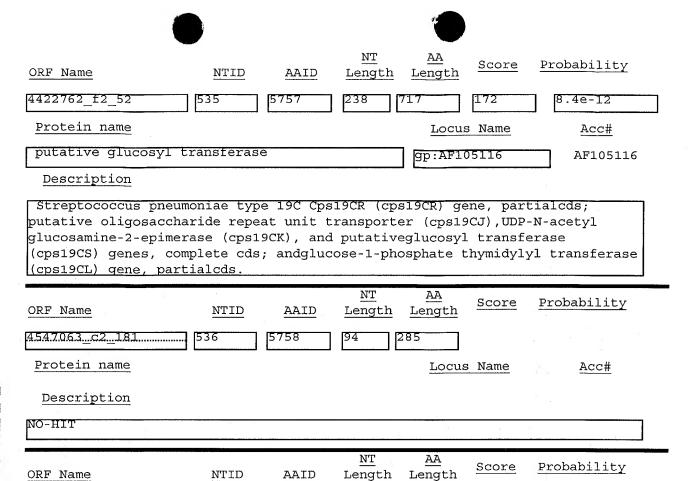
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
24713961_f2_37	523	5745	304	915	376	1.3e-34
<u>Protein name</u>				Locus	Name	Acc#
prolipoprotein d (lgt) RP046	iacylglyceryl	transfe	rase	pir:F7	712	F71712
Description				•		
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
25505386 <u>f14</u>	524	5746	244	735	232	2.3e-19
Protein name				Locus	Name	Acc#
chloramphenicol	acetyltransfe	rase		gp:AF12	4757	AF124757
Description	,					x e = *
Zymomonas mobili	s fosmid clon	ne 43D2, d	complete	sequence		
	-		NT	AA		m 1 1 1 1 1 1
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 2557938382	***************************************	<u>AAID</u> 5747	Length		Score 81	Probability
*	***************************************		Length	Length	-	
255.7.93.83 <u></u> £3 <u></u> 82	***************************************		Length	Length	81 Name	0.0020
255.7.93.83 <u></u> £3 <u></u> 82	***************************************		Length	Length  884  Locus	81 Name	0.0020 <u>Acc#</u>
25579383_f3_82 Protein name	525	5747	Length	Length  884  Locus	81 Name	0.0020 <u>Acc#</u>
25579383_f3_82  Protein name  Description	525	5747	Length 627	Length  884  Locus  sp:EREE	81 Name	0.0020 <u>Acc#</u>
25579383_f3_82  Protein name  Description	525	5747	Length	Length  884  Locus	81 Name	0.0020 <u>Acc#</u>
25579383_f3_82  Protein name  Description  ERYTHROMYCIN EST	ERASE TYPE II	5747	Length 627 [1	Length  B84  Locus  Sp:EREE  AA  Length	81 Name B_ECOLI	0.0020 <u>Acc#</u> P05789
255.793.83	ERASE TYPE II	5747 ,	Length  627  NT  Length	Length  Locus  Sp:EREE  AA  Length  125	81 Name B_ECOLI Score	0.0020 Acc# P05789  Probability
255.793.83_f3_82  Protein name  Description  ERYTHROMYCIN EST  ORF Name  263.6.71.35_f3_70	ERASE TYPE II	5747 ,	Length  627  NT  Length	Length  Locus  Sp:EREE  AA  Length  125	81  Name  Score  1148  Name	0.0020 Acc# P05789  Probability 2.0e-116
255.793.83_f3_82  Protein name  Description  ERYTHROMYCIN EST  ORF Name  263.6.71.35_f3_70	ERASE TYPE II	5747 ,	Length  627  NT  Length	Length  Locus  Sp:EREE  AA  Length  125  Locus	81  Name  Score  1148  Name	0.0020  Acc#  P05789  Probability  2.0e-116  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
26757637_f3_88	527	5749		2208	867	1.2e-86	_
		3,43	ا الـــــــــــــــــــــــــــــــــــ			J [	
Protein name				Locus	s Name	Acc#	
hemolysin secretic sll1180:protein sll	_	nlyB:prote	ein	pir:S7	5806	S75806	
Description				•			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
2741426 <u></u> £3 <u></u> 98	528	5750	397 1	194	355	2.1e-32	
Protein name				Locus	Name	Acc#	
				sp:PBP_	BACSU	P39844	
Description				,		- ÷	
PUTATIVE PENICILLI	N BINDING	PROTEIN PR	RECURSOR				
	-		NT	AA			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	-
ORF Name 2928307c3215		<u>AAID</u> 5751	Length		<u>Score</u>	Probability  2.4e-120	
			Length	Length		1	
			Length	Length	1185 Name	2.4e-120	
			Length	Length 584 Locus	1185 Name	2.4e-120	
2928307 <u></u> c3 <u></u> 215 Protein name	529		Length 527	Length 584 Locus	1185 Name	2.4e-120	
Protein name  Description	529	5751	Length  [527]  [1]  THETASE E	Length  .584  Locus  sp:NADE	1185 Name	2.4e-120 Acc#	
Protein name  Description	529	5751	Length 527	Length 584 Locus	1185 Name	2.4e-120	
2928307_c3_215	EE, (QUINOL	5751 INATE SYNT	Length  527  THETASE B  NT  Length	Length  Locus  sp:NADE	1185  Name  B_PSEAE	2.4e-120 Acc#	
2928307_c3_215	EE, (QUINOL:	5751 INATE SYNT	Length  527  THETASE B  NT  Length	Length Locus Sp:NADE	1185 S Name  B PSEAE  Score	2.4e-120 Acc# Probability	
Protein name  Description  L-ASPARTATE OXIDAS  ORF Name  33992307_t1_23	EE, (QUINOL:	5751 INATE SYNT	Length  527  THETASE B  NT  Length	Length Locus Sp:NADE	I185 S Name SCORE  257 S Name	Probability  [5.1e-22]	
Protein name  Description  L-ASPARTATE OXIDAS  ORF Name  33992307_t1_23	EE, (QUINOL:	5751 INATE SYNT	Length  527  THETASE B  NT  Length	Length  Locus  Sp:NADE  AA  Length  Locus  Locus	I185 S Name SCORE  257 S Name	Probability  5.1e-22  Acc#	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
	<del></del>				D C 0 2	1 1 00 164	_,
34376678_c1_141	531	5753	892	:679	1603	1.2e-164	╛
Protein name				Locus	s Name	Acc#	
				sp:MUT	S_HAEIN	P44834	
Description	* a						
DNA MISMATCH REPAI	R PROTEIN	MUTS					
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
4094128_f2_51	532	5754	599 1	800	83	0.026	
Protein name				Locus	s Name	Acc#	
erythromycin ester	ase homolog	g ybf0		pir:A6	9750	A69750	
Description	V-1				*	•	
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability	
424042 <u></u> £3 <u></u> 103	533	5755	470 1	413	557	1.2e-101	
Protein name				Locus	s Name	Acc#	
putative protein				gp:ATA	P22	Z99708	
Description			-		*	16 16	
Arabidopsis thalia	na DNA chro	omosome 4,	ESSA I	AP2 cont	ig fragm	mentNo. 2.	
		*	NT	7.7	-		
ORF Name	NTID	AAID	Length	<u>AA</u> Length	Score	Probability	
4332837 <u></u> £3 <u></u> 86	534	5756	228 6	87			
Protein name				Locus	s Name	Acc#	
Description						*	
NO-HIT	· · · · · · · · · · · · · · · · · · ·				7		

4569126<u>\_\_</u>£2<u>\_\_</u>68\_\_\_\_

Protein name



sp:PBP4\_HAEIN P45161 Description NT AAScore Probability ORF Name NTID AAID Length Length 4901587\_c1\_114 538 5760 562 1689 1101 1.9e-111 Protein name Locus Name Acc# probable sulfate transporter pir:A71463 A71463 Description

88

267

Locus Name

5759

537

0.0057

Acc#

ODE Name	NTID AAID	NT AA Score Probability Length Length	
ORF Name			_
5985875_c3_220	539 5761	810 2433 615 5.9e-60	ا
Protein name		Locus Name Acc#	
ferrichrome-iron r	_	pir:S74457 S74457	
slr1490:protein slr			
Description			
		NT AA Court Buckshildt	
ORF Name	NTID AAID	Length Length Score Probability	
6767057 <u>t2</u> 34	540 5762	3.7e-09	
Protein name		Locus Name Acc#	
hypothetical prote	ein PAB1767	pir:B75136 B75136	
Description			
ORF Name	NTID AAID	NT AA Score Probability	
813302 £1 9	541   5763	373 1122	
		Locus Name Acc#	
<u>Protein name</u>		nocus name	
Description			
NO-HIT			
	-	NT AA Grove Drobability	V 6
ORF Name	NTID AAID	Length Length Score Probability	
970680f110	542 5764	965 2898 1588 5.1e-217	
Protein name		Locus Name Acc#	
putative leucyl th	RNA synthetase	gp:AF069441 AF069441	
Description	-		
Arabidopsis thalia	ana BAC T19B17 from	chromosome IV, near 19.3 cM, complete	
sequence.			

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
10158452 c2 398	543	5765		753	587	5.5e-57
Protein name				_	s Name	Acc#
putative glycosyl	transierase	<u> </u>		gp:AF0	48749	AF048749
Description						
Bacteroides fragi sequence.	lis capsula:	r polysac	charide .	biosynthe	esis ope	ron, complete
			NT	AA	-	
ORF Name	NTID	AAID	Length	Length	Score	Probability
1054637_f3_214	544	5766	211	636	1033	3.0e-104
Protein name				Locus	s Name	Acc#
superoxide dismut	ase			gp:BNR	SOD2	D13756
Description						
Bacteroides fragi	lis DNA for	superoxio	de dismu	tase, com	mplete co	ds.
		-		•	-	The state of the s
		-				
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
ORF Name  10.7.5.0.0.6.7			NT Length	AA		
		AAID	NT Length	AA Length		
		AAID	NT Length	AA Length	Score	Probability
10750067c2445 Protein name		AAID	NT Length	AA Length	Score	Probability
Protein name  Description		AAID	NT Length	AA Length 381 Locus	Score S Name	Probability  Acc#
Protein name  Description		AAID	NT Length	AA Length	Score	Probability
Protein name  Description  NO-HIT	545 NTID	<u>AAID</u> 5767	NT Length 126	AA Length 381 Locus	Score S Name	Probability  Acc#
Description NO-HIT ORF Name	545 NTID	AAID 5767  AAID	NT Length 126	AA Length  Locus  AA Length  AA Length	Score Score	Probability  Acc#  Probability
10.750.06.7	NTID	AAID 5767  AAID	NT Length 126	AA Length  Locus  AA Length  AA Length	Score  Score  Score  803  Name	Probability  Acc#  Probability  7.1e-80
Description NO-HIT  ORF Name  10.803580_c2_394  Protein name	NTID	AAID 5767  AAID	NT Length 126	AA Length  Locus  AA Length  AA Length  Locus  Locus	Score  Score  Score  803  Name	Probability  Acc#  Probability  7.1e-80  Acc#

ORF Name	NTID AAID Le	NT AA score	Probability
10837887_c1_303  Protein name	547 5769 37	11125 1002 Locus Name	5.8e-101 Acc#
CDP-glucose-4,6-	lehydratase	pir:D47070	D47070
ORF Name 10978425c3473	NTID AAID Le	NT AA Score	Probability
Protein name		Locus Name	Acc#
Description NO-HIT			
ORF Name		NT <u>AA</u> ngth <u>Length</u> <u>Score</u>	Probability
11214032_c2_3.77	549 5771 47		1.8e-83
Protein name		Locus Name sp:ATOC_ECOLI	<u>Acc#</u> Q06065
Description DECARBOXYLASE IN	HIBITOR) (ORNITHINE DECAR	RBOXYLASE ANTIZYME)	
ORF Name		NT AA ngth Length Score	Probability
11932290 <u></u> £2 <u></u> 88	550 5772 10	7 324 152	1.2e-10
Protein name		Locus Name	Acc#
		sp:CBIK SALTY	Q05592
Description  CBIK PROTEIN		SP.CBIK_SAUIT	Q05592

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
126376 f3 197	551	5773		503		
	331	3,73				
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
13.785926 <u></u> c2 <u></u> 3.72	552	5774	172	519		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						-
			NT	AA		
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability
13834812 <u></u> c3 <u></u> 485	553	5775	347	.044		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA		
ORF Name	NTID	AAID	Length	<u>AA</u> Length	Score	Probability
14461567 <u></u> £3 <u></u> 222	554	5776	386	161	890	4.3e-89
Protein name				Locus	Name	Acc#
ThiH				gp:AF15	4064	AF154064
Description						
Salmonella typhimur	ium ThiH	(thiH) ger	ne, compl	lete cds.		

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
14489050_£2_180	555 5	777	555 1	.668	147	4.7e-07
Protein name				Locus	s Name	Acc#
aspartate aminotrar	nsferase			pir:D7	5496	D75496
Description						
			NT	AA	Caoro	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability
14642207 <u></u> ±1 <u></u> 18	556 5	778	595 1	.788	1878	8.6e-194
Protein name				Locus	s Name	Acc#
				sp:THI	C_BACSU	
Description						
THIAMINE BIOSYNTHES	SIS PROTEIN	THIC				* 10
ODE Name	NETE	3 3 TD	NT	AA Taranti	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>		
14647206 <u></u> £3 <u></u> 237		<u>AAID</u> 779	Length	Length 74	96	0.021
14647206f3237 Protein name	557 [5	779	Length	Length 74 Locus	96 Name	0.021 Acc#
Protein name  conserved hypothetic	557 [5	779	Length	Length 74	96 Name	0.021
14647206f3237 Protein name	557 [5	779	Length	Length 74 Locus	96 Name	0.021 Acc#
Protein name  conserved hypothetic	557 [5	779	Length	Length 74 Locus	96 Name	0.021 Acc#
Protein name  conserved hypothetic  Description	cal protein	779 MTH469	Length  257  7  NT  Length	Length  74  Locus  pir:D69	96 S Name 9161	0.021 Acc# D69161
Protein name  conserved hypothetic  Description  ORF Name	cal protein	779 MTH469	Length  257  7  NT  Length	Length  Locus  pir:D69  AA  Length	96 Name 9161 Score	D69161  Probability
Protein name  Conserved hypothetic  Description  ORF Name  14729186t2148	cal protein	779 MTH469	Length  257  7  NT  Length	Length  Locus  pir:D69  AA  Length	96 S Name 9161 Score 88 Name	D69161  Probability  0.029
Protein name  Conserved hypothetic  Description  ORF Name  14729186t2148	cal protein	779 MTH469	Length  257  7  NT  Length	Length  Locus  pir:D69  AA  Length  68  Locus	96 S Name 9161 Score 88 Name	D69161  Probability  0.029  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
14867327_f3_236	559	5781	408	1227	175	3.0e-10
Protein name				Locus	s Name	Acc#
				sp:YIG	M_ECOLI	P27850
Description				<u> </u>		
HYPOTHETICAL 54.7	KD PROTEIN	IN UDP-U	BIE INTE	RGENIC RI	EGION PRI	ECURSOR
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
16064015_c2_385	560	5782	140	423	84	0.0060
Protein name				Locus	s Name	Acc#
trbA protein				pir:A4	9852	
Description					*	-
7				707		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
164043 <u></u> c1 <u></u> 309	561	5783	251	756	493	5.0e-47
Protein name				Locus	s Name	Acc#
conserved hypothe	tical prote	in HP0162		pir:B6	4540	B64540
Description				-		
*				*		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
	MIID	MIL		<u> </u>		
194086c3520		5784		2373	1337	2.8e-144
194086c3520 Protein name				2373	1337 S Name	2.8e-144 Acc#
				2373 Locus		
				2373 Locus	s Name	Acc#

		NT	AA		
ORF Name	NTID A	AID Length	***************************************	Score	Probability
1960876_c2_403	563 578	5 72	219	81	0.0023
Protein name			Locus	s Name	Acc#
hypothetical prote	in MJ1608		pir:G6	4500	G64500
Description					
		2777	73. 78		
ORF Name	NTID A	<u>NT</u> AID <u>Length</u>	<u>AA</u> Length	Score	Probability
19689678 <u></u> c1 <u></u> 307	564 578	6 442	1329	128	5.0e-05
Protein name			Locus	s Name	Acc#
unknown			gp:AF14	14879	AF144879
Description					
Leptospira interro	gans rfb locus	, complete se	equence.	× × × × × × × × × × × × × × × × × × ×	
		NT	AA		
ORF Name	NTID A	AID Length	<del></del>	Score	Probability
19709682 <u>c1</u> 302	565 578	7 451	1356	1278	3.3e-130
Protein name			Locus	Name	Acc#
CDP-4-keto-6-deoxy	-D-glucose-3-de	hydratase	gp:YPE2	251713	AJ251713
Description	, v				-
Yersinia pestis st for ddhD gene, ddhA wzx gene, wbyI pseu pseudogene, fcl pse gene (partial).	A gene, ddhB psa adogene, wbyJ g	eudogene, ddł ene, wzypseud	nC gene, p dogene, wb	ortgene, oyK gene,	wbyH gene, gmd
ORF Name	NTID A	<u>NT</u> AID Length	<u>AA</u> Length	Score	Probability
19711067 c3 486	-		903	374	2.1e-34
		, 300	8		
Protein name				Name ·	Acc#
hypothetical prote	in jhp0094		pir:E71	.975	E71975
Description					

NTID   AAID   Length   Lengt			$\underline{ ext{NT}}$	AA Sco	re Probability
Protein name	ORF Name	NTID AAID	Length	Length Soon	11000011110
Putative UDP-GICNAc:undecaprenylphosphate   Ep:AF048749   AF048749	20087751_c3_489	567 5789	318	1429	3.3e-146
Description	Protein name			Locus Nam	e <u>Acc#</u>
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.	putative UDP-GlcN	Ac:undecaprenylphosp	hate	gp:AF048749	AF048749
Sequence	Description				
ORF Name         NTID         AAID         Length         Length         Score         Probability           20520302_c3_462         568         5790         495         1488         112         9.6e-06           Protein name         Locus Name         Acc#           immunoreactive 50kD antigen PG53         gp:AF175720         AF175720           Description         Porphyromonas gingivalis strain W50 immunoreactive 50kD antigenPG53 gene, complete cds.         NTID         AAID         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         Locus Name         Acc#           Terrichrome-iron receptor 3:protein slr1490         pir:S74457         S74457           Description         NTID         AAID         Length         Length         Score         Probability           Protein name         NTID         AAID         Length         Length         Acc#           Protein name         Locus Name         Acc#           Description         Locus Name         Acc#		lis capsular polysac	charide b	piosynthesis	operon, complete
Description			NT	AA Saa	ro Drobobility
Description	ORF Name	NTID AAID	Length	Length Sco	re Probability
Description	20520302_c3_462	568 5790	495 1	488 112	9.6e-06
Description	Protein name			Locus Name	e <u>Acc#</u>
Porphyromonas gingivalis strain W50 immunoreactive 50kD antigenPG53 gene, complete cds.   NTD	immunoreactive 50	kD antigen PG53		gp:AF175720	AF175720
ORF Name         NTID         AAID         Length         Length         Score         Probability           20596012_f1_1         569         5791         801         2406         I81         9.9e-33           Protein name         Locus Name         Acc#           ferrichrome-iron receptor 3:protein slr1490         pir:S74457         S74457           Description         NTID         AAID         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Probability           21495928_f1_23         570         5792         94         285           Protein name         Locus Name         Acc#           Description	Description		0.		
ORF Name         NTID         AAID         Length         Length         Score         Probability           20596012_f1_1_         569         5791         801         2406         181         9.9e-33           Protein name         Locus Name         Acc#           ferrichrome-iron receptor 3:protein slr1490         pir:S74457         S74457           Description         NTID         AAID         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Probability           21495928_f1_23         570         5792         94         285           Protein name         Locus Name         Acc#           Description         Description		givalis strain W50 i	mmunoreac	tive 50kD an	tigenPG53 gene,
ORF Name         NTID         AAID         Length         Length         Length           20596012_f1_1         569         5791         801         2406         181         9.9e-33           Protein name         Locus Name         Acc#           ferrichrome-iron receptor 3:protein slr1490         pir:S74457         S74457           Description         NTID         AAID         NTID         AA           ORF Name         NTID         AAID         Length         Score         Probability           21495928_f1_23         570         5792         94         285           Protein name         Locus Name         Acc#           Description         Locus Name         Acc#		41	NT	AA SCO	re Probability
Protein name	ORF Name	NTID AAID	<u>Length</u>	Length Sco.	riobability
Description	20596012 <u></u> f1 <u></u> 1	569 5791	801 2	406 181	9.9e-33
Description	Protein name			Locus Name	<u>Acc#</u>
Description         NT         AA         Score         Probability           ORF Name         NTID         AAID         Length         Length         Score         Probability           21495928±123				pir:S74457	S74457
ORF Name         NTID         AAID         Length         Length         Score         Probability           21495928±123		11490	-		
ORF Name         NTID         AAID         Length         Score         Probability           21495928t123	Description				
ORF Name         NTID         AAID         Length         Length           21495928t1_23			NT	AA a	7 1 1 17 1
Protein name Locus Name Acc#  Description	ORF Name	NTID AAID	Length	Length Scor	re Probability
Description	21495928 <u></u> t1 <u></u> 23	570 5792	94 2	85	
	Protein name			Locus Name	<u>Acc#</u>
NO-HIT	Description				4
	NO-HIT				

			$\underline{ ext{NT}}$	$\underline{\mathtt{A}\mathtt{A}}$	Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	30010	
2150305_c1_308	571	5793	299	900	174	1.1e-20
Protein name				Locus	s Name	Acc#
UDP-glucose-4-epi	merase/dTDP-	glucose-4	, 6	gp:AF04	18749	AF048749
Description				1		
Bacteroides fragi sequence.	lis capsular	polysaco	charide b	iosynthe	esis oper	con, complete
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
22114755_f1_7	572	5794	478 1	.437	384	1.8e-35
Protein name				Locus	Name	Acc#
precorrin-6Y methy sll0099:protein sl		n		pir:S76	697	S76697
Description					9	e e
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
22129152 <u></u> £2 <u></u> 112	573	5795	632 1	.899		
Protein name				Locus	Name	Acc#
Description						
NO-HIT			W 1.77 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	<b></b>	-	, x
			D.T.T.	7.7		1
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2349150 <u></u> c1 <u></u> 344	574	5796	103	12		
Protein name				Locus	Name	Acc#
Description						•
NO-HIT						

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2351432_f2_175  Protein name	575	5797	68	Locus	s Name	Acc#
Description	· · · · · · · · · · · · · · · · · · ·					
LVO-1111						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
23.56.2.762 <u></u> c3 <u></u> 488	. 576	5798	240	723	366	1.4e-33
Protein name				Locus	s Name	Acc#
putative glycosyl	transferas	e ·		gp:AF0	71085	AF071085
				J [		
Description						
Description Enterococcus faeca partial sequence.	lis strain	OG1RF po	141		yntheti:	c genecluster,
Enterococcus faeca	lis strain	OGIRF po	lysaccha <u>NT</u> Length	ride bios <u>AA</u> Length	Score	Probability
Enterococcus faeca partial sequence.	NTID		NT Length	AA		
Enterococcus faeca partial sequence. ORF Name	NTID	AAID	NT Length	<u>AA</u> Length		
Enterococcus faeca partial sequence.  ORF Name  23611383t138	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
Enterococcus faeca partial sequence.  ORF Name  23611383_f1_38	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
Enterococcus faeca partial sequence.  ORF Name  23611383_f1_38	NTID	AAID	NT Length	AA Length 207 Locus	Score S Name	Probability Acc#
Enterococcus faeca partial sequence.  ORF Name  23611383_f1_38	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
Enterococcus faeca partial sequence.  ORF Name  23611383_f1_38  Protein name  Description  NO-HIT	NTID    577	AAID 5799	NT Length 68	AA Length 207 Locus	Score S Name	Probability  Acc#
Enterococcus faeca partial sequence.  ORF Name  23611383tl38  Protein name  Description  NO-HIT  ORF Name	NTID    577	<u>AAID</u> [5799]	NT Length 68	AA Length  Locus  AA Length  1230	Score Score	Probability  Acc#  Probability
Enterococcus faeca partial sequence.  ORF Name  23.6.113.83	NTID    577	<u>AAID</u> [5799]	NT Length 68	AA Length  Locus  AA Length  1230	Score Score 327 Name	Probability  Acc#  Probability  6.3e-32
Enterococcus faeca partial sequence.  ORF Name  23.6.113.83	NTID    577	<u>AAID</u> [5799]	NT Length 68	AA Length  Locus  AA Length  Locus  Locus  Locus  Locus	Score Score 327 Name	Probability  Acc#  Probability  6.3e-32  Acc#

Second Name	ORF Name	NTID AAID	NT AA Score Pr Length Length	obability
PF31 39K orf136	23617802_f2_160	579 5801	82 249 84	0.0054
Description	Protein name		Locus Name	Acc#
ORF Name         NTID         AAID         Length         Length         Score         Probability           23.7.04658_c3_459         580         5802         439         1320         132         1.8e-11           Protein name         Locus Name         Acc#           conserved hypothetical protein yknZ         pir:E69858         E69858           Description         NTID         AAID         NT         AA         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Acc#         Acc#           Description         NO-HIT         AAID         Length         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Acc#         Acc#           Protein name         Locus Name         Acc#	PP31 39K ort36		pir:T41782	T41782
ORF Name         NTID         AAID         Length         Length         Secret         Frobatity           23.704688_c1_459         580         5802         439         1320         132         1.8e-11           Protein name         Locus Name         Acc#           Conserved hypothetical protein yknZ         pir:E69858         E69858           Description         NTID         AAID         Length         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Length         Acc#           Description         NO-HIT         AAID         Length         Length         Score         Probability           NO-HIT         Description         NO-HIT         AAID         Length         Length         Acc#           Description         NO-HIT         AAID         Length         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Acc#           Description         NO-HIT         AAID         Length         Length         Length           Description         NO-HIT         AAID         Length         Length         Length         Acc#	Description			
ORF Name         NTID         AAID         Length         Length         Secret         Frobatity           23.704688_c1_459         580         5802         439         1320         132         1.8e-11           Protein name         Locus Name         Acc#           Conserved hypothetical protein yknZ         pir:E69858         E69858           Description         NTID         AAID         Length         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Length         Acc#           Description         NO-HIT         AAID         Length         Length         Score         Probability           NO-HIT         Description         NO-HIT         AAID         Length         Length         Acc#           Description         NO-HIT         AAID         Length         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Acc#           Description         NO-HIT         AAID         Length         Length         Length           Description         NO-HIT         AAID         Length         Length         Length         Acc#			NT AA	-1-1-1-1-1-
Protein name	ORF Name	NTID AAID	Score Pr	Obability
Description	23.7046.88 <u></u> c3 <u></u> 45.9	580 5802	1320 132	1.8e-11
Description   Description   NTID   AAID   Length   Leng	Protein name		Locus Name	Acc#
ORF Name         NTID         AAID         Length         Length         Score         Probability           23831325_f2_155         581         5803         95         288           Protein name         Locus Name         Acc#           Description         NO-HIT         NT         AA         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         Locus Name         Acc#           Description           NO-HIT           ORF Name         NTID         AAID         NT         AA         Score         Probability           23851637_f3_217         583         5805         266         801         824         4.2e-82           Protein name         Locus Name         Acc#           Description	conserved hypoth	etical protein yknZ	pir:E69858	E69858
ORF Name         NTID         AAID         Length         Length         Score         Frobability           Protein name         Locus Name         Acc#           Description           NO-HIT           ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         Locus Name         Acc#           Description           NO-HIT         NTID         AAID         Length         Length         Acc#           Protein name         NTID         AAID         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         4.2e-82           Protein name         Locus Name         Acc#           Description	Description			
NTID   AAID   Length   Length	<u> </u>		NT AA Score Pr	robability
Protein name	ORF Name	NTID AAID	Length Length	
Description   No-HIT	23.83.13.25 <u></u> £2 <u></u> 15.5	581 5803	95 288	
No-HIT	Protein name		Locus Name	Acc#
ORF Name         NTID         AAID         Length         Length         Score         Probability           23851567‡2119				
ORF Name         NTID         AAID         Length         Length         Score         Probability           23851567f2119	Description			**
ORF Name         NTID         AAID         Length         Length           2385156.7f2119         582         5804         192         579           Protein name         Locus Name         Acc#           Description           NO-HIT         NTID         AAID         NTID         AAID         Length         Score         Probability           23851637f3217         583         5805         266         801         824         4.2e-82           Protein name         Locus Name         Acc#           Sp:THIG_ECOLI         Sp:THIG_ECOLI				* * *
Protein name         Locus Name         Acc#           Description         NO-HIT         NT AA Score         Probability           ORF Name         NTID AAID Length Length         Score Probability           23851637_f3_217583         5805         266         801         824         4.2e-82           Protein name         Locus Name         Acc#           Description         Sp:THIG_ECOLI			NT AA Score Pi	
Description   NO-HIT   NT   AA   Score   Probability	NO-HIT	NTID AAID	— Score Pi	
NO-HIT         NT         AA         Score         Probability           23851637±3217	NO-HIT ORF Name		Length Length	
ORF Name         NTID         AAID         Length         Length         Score         Probability           23851637±3217	NO-HIT  ORF Name  23851567_f2_119		Length Length Score F1	cobability
ORF Name         NTID         AAID         Length         Length         Floatility           23851637_f3_217	ORF Name  23851567_f2_119  Protein name		Length Length Score F1	cobability
ORF Name         NTID         AAID         Length         Length           23851637_f3_217	ORF Name  23851567_f2_119  Protein name  Description		Length Length Score F1	cobability
Protein name  Locus Name  Acc#  sp:THIG_ECOLI  Description	ORF Name  23851567_f2_119  Protein name  Description		Length Length  192   579    Locus Name	Acc#
protein name  sp:THIG_ECOLI  Description	NO-HIT  ORF Name  23851567f2119  Protein name  Description  NO-HIT	[582] [5804]	Length Length  192   579  Locus Name  NT AA Length Length Score Processing Pr	Acc#
Description	NO-HIT  ORF Name  23851567f2119  Protein name  Description  NO-HIT  ORF Name	NTID AAID	Length Length Score Process Name    192   579	Acc# robability 4.2e-82
	NO-HIT  ORF Name  23851567_f2_119  Protein name  Description  NO-HIT  ORF Name  23851637_f3_217	NTID AAID	Length Length  192   579   Locus Name    NT	Acc# robability 4.2e-82
THIG PROTEIN	NO-HIT  ORF Name  23851567_f2_119  Protein name  Description  NO-HIT  ORF Name  23851637_f3_217	NTID AAID	Length Length  192   579   Locus Name    NT	Acc# robability 4.2e-82
	ORF Name  23851567f2119  Protein name  Description  NO-HIT  ORF Name  23851637f3217  Protein name	NTID AAID	Length Length  192   579   Locus Name    NT	Acc# robability 4.2e-82

		NT AA Score Probability
ORF Name	NTID AAID	Length Length
23957812_f3_267	584 5806	109 330 91 0.0011
Protein name		Locus Name Acc#
chaperone GrpE ty	pe 2	gp:AF098636 AF098636
Description		
	n chaperone GrpE type tein, complete cds.	2 (GrpE2) mRNA, nuclear geneencoding
1		NT AA Score Probability
ORF Name	NTID AAID	Length Length
24023462_c1_311	585 [5807	426 1281 618 2.9e-60
Protein name		Locus Name Acc#
		sp:YDAR_BACSU P96593
Description		
HYPOTHETICAL 45.	7 KD PROTEIN IN MUTT-	GSIB INTERGENIC REGION
		NT AA Duchahilitaa
ORF Name	NTID AAID	Length Length Score Probability
24035952 <u>f2147</u>	586 5808	95 288 82 0.0018
Protein name		Locus Name Acc#
unknown protein		gp:SCCXV106K
Description		
S.cerevisiae 10.	6kbp fragment from ch	romosome XV.
		NT AA
ORF_Name	NTID AAID	$rac{ ext{NT}}{ ext{Length}}$ $rac{ ext{AA}}{ ext{Length}}$ Score Probability
24223762_c1_318	587 5809	448 1347 503 1.6e-82
Protein name		Locus Name Acc#
Na+/H+-exchanging	g protein:Na+/H+ anti	porter pir:JX0360 JX0360
Na+/H+-exchanging Description	g protein:Na+/H+ anti	porter pir:JX0360 JX0360

		N	. AA		D 1 1 1 1 1 1 1
ORF Name	NTID	AAID Leng	th Length	Score	Probability
24239006_£3_269	588 58	337	1014	110	0.0058
Protein name			Locus	<u>Name</u>	Acc#
			gp:ECO	RHSEX	L19083
Description					*
Escherichia coli					otein, complete
cds; complete ORF-1	E2; H-rpt sube	element; com	plete ORF-H.		
ODE N	NULL	NT Long	-	Score	Probability
ORF Name	NTID	AAID Leng			
24303127_f2_173	589 58	142	429		
Protein name			Locus	s Name	Acc#
Description					
NO-HIT					
				14	
		NIG	7.7.		
ORF Name	NTID	NT AAID Leng		Score	Probability
ORF Name 24407687				<u>Score</u>	Probability  1.1e-120
		AAID Leng	Length 2484		
24407687 <u></u> c2 <u></u> 400		AAID Leng	Length 2484	1188 Name	1.1e-120
24407687 <u></u> c2 <u></u> 400		AAID Leng	Locus	1188 Name	1.1e-120
24407687c2400 Protein name	<u>590</u> <u>58</u>	AAID Leng	Locus	1188 Name	1.1e-120
Protein name  Description	<u>590</u> <u>58</u>	AAID Leng	th Length  2484  Locus  sp:SYFF	1188 Name	1.1e-120
Protein name  Description	<u>590</u> <u>58</u>	AAID Leng	Locus  sp:SYFI	1188 Name	1.1e-120
Protein name  Description  TRNA LIGASE BETA	EHAIN) (PHERS)	AAID Leng	Locus  sp:SYFI	1188  Name  B_ECOLI	1.1e-120 Acc#
Protein name  Description  TRNA LIGASE BETA CONTRACTORY	EHAIN) (PHERS)	AAID Leng NT AAID Leng	Length  Locus  Sp:SYFI  AA th Length	1188  Name  B_ECOLI	1.1e-120 Acc#
Protein name  Description  TRNA LIGASE BETA (  ORF Name  24410780t171	EHAIN) (PHERS)	AAID Leng NT AAID Leng	Length  Locus  Sp:SYFI  AA th Length	1188 S Name S ECOLI Score	Acc#  Probability  Acc#
Protein name  Description  TRNA LIGASE BETA CONTROL NAME  ORF Name	EHAIN) (PHERS)	AAID Leng NT AAID Leng	Length  Locus  Sp:SYFI  AA th Length	1188 S Name S ECOLI Score	1.1e-120 Acc#  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24412912_f2_94	592	5814	192 5	79			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA		- 1 1171	*
ORF Name	NTID	AAID	Length	Length	Score	Probability	
24504713 <u></u> ±3200	593	5815	240 7	/23	291	1.3e-25	
Protein name				Locus	Name	Acc#	
hypothetical pro	tein MTH671			pir:D69	189	D69189	
Description						-	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 24614415t133		<u>AAID</u> 5816	Length		Score	Probability	
-			Length	Length			
24614415 <u></u> £1 <u></u> 33			Length	Length	337 Name	1.7e-30	
24614415 <u></u> £1 <u></u> 33			Length	Length  O48  Locus	337 Name	1.7e-30	
24614415t133 Protein name	594	5816	Length 315	Length  Assume Locus  Sp:YLYE	337 Name BACSU	1.7e-30 Acc#	- ].   .
Protein name  Description	594	5816	Length 315 S	Length  Locus  Sp:YLYE	337 Name BACSU	1.7e-30 <u>Acc#</u> RF-X)	
Protein name  Description	594	5816	Length 315	Length  Assume Locus  Sp:YLYE	337 Name BACSU	1.7e-30 Acc#	
Protein name  Description  HYPOTHETICAL 33		IN LSP-P	Length  315  YRR INTER  NT  Length	Length  Locus  Sp:YLYE	337 Name BACSU GION (OF	1.7e-30 <u>Acc#</u> RF-X)	
Protein name  Description  HYPOTHETICAL 33		IN LSP-P	Length  315  YRR INTER  NT  Length	Length  Locus  Sp:YLYE  AA  Length  Length	337 Name BACSU GION (OF	1.7e-30 Acc#  RF-X)  Probability	
Protein name  Description  HYPOTHETICAL 33.  ORF Name  24641937tl8.	.7 KD PROTEIN  NTID  595	IN LSP-P	Length  315  YRR INTER  NT  Length	Length  Locus  Sp:YLYE  AA  Length  Length	337 Name BACSU GION (OF Score 628 Name	1.7e-30 Acc#  RF-X)  Probability  2.5e-61	
Protein name  Description  HYPOTHETICAL 33  ORF Name  24641937_t1_8  Protein name	.7 KD PROTEIN  NTID  595	IN LSP-P	Length  315  YRR INTER  NT  Length	Length  Locus  Sp:YLYE  RGENIC RE  AA  Length  .806	337 Name BACSU GION (OF Score 628 Name	1.7e-30 Acc#  RF-X)  Probability  2.5e-61 Acc#	

F.	
1	120
1	M
	::25
:;;	131
1	1174 1187
ķ	1135
-	
15	
1,165	125
250	
	1371
ľ	Had
	983 H3G
ľ	983 H3G
ľ	120 125 126 126
**************************************	
**** * **** **	
**************************************	
**** * **** **	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24642311_c2_392	596	5818	162 4	.89 Locus	223 Name	2.1e-18 Acc#
Protein name unknown				gp:AF04		AF048749
Description						
Bacteroides fragil sequence.	lis capsular	polysaco	charide b	oiosynthe	sis ope	con, complete
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24648563_c1_322	597	5819	136 4	:11		
Protein name				Locus	Name	Acc#
Description				=		
NO-HIT						
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
		<u>AAID</u> [5820	Length		Score 659	Probability 1.8e-115
ORF Name			Length	Length		
ORF Name  246.94013f13	598	5820	Length	Length	659 Name	1.8e-115
ORF Name  246.940.13f13  Protein name	598	5820	Length	Length  3981  Locus	659 Name	1.8e-115 <u>Acc#</u>
ORF Name  246.94013f13  Protein name  cobalamin biosynth	598	5820	Length	Length  3981  Locus	659 Name	1.8e-115 <u>Acc#</u>
ORF Name  246.94013_f1_3  Protein name  Cobalamin biosynth  Description	nesis protei	5820 In N	Length  1326  NT Length	Length  1981  Locus  pir:C65	659 S Name	1.8e-115 Acc# C69048
ORF Name  246.94013_f1_3  Protein name  Cobalamin biosynth  Description  ORF Name	nesis protei	En N	Length  1326  NT Length	Length  Locus  Pir:C69  AA  Length	659 S Name 9048 Score	1.8e-115  Acc#  C69048  Probability
ORF Name  246.94013fl3  Protein name  Cobalamin biosynth  Description  ORF Name  25906675c3492	NTID	En N	Length  1326  NT Length	Length  Locus  Pir:C69  AA  Length	Score 106 Name	1.8e-115 Acc# C69048  Probability  3.1e-05

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
25976708_f3_209	600	5822	95	288		
Protein name				Locus	Name	Acc#
Description						
NO-HIT					-	
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
26362791 <u>f3</u> 272	601	5823	603	.812	683	3.7e-67
Protein name				Locus	Name	Acc#
probable membrane p	rotein b0	847		pir:G64	822	G64822
Description	-					*
			2777	7.7		<u> </u>
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
26604510 <u></u> 314	602	5824	595	.788	1902	2.5e-196
Protein name				Locus	Name	Acc#
				sp:LEPA	_BACSU	P37949
Description						· · ·
GTP-BINDING PROTEIN	N LEPA					
		14	NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
26835887 <u></u> t1 <u></u> 20	603	5825	238	17	518	1.1e-49
Protein name		=		Locus	Name	Acc#
MPT-synthase sulfur	ylase	· · · · · · · · · · · · · · · · · · ·		gp:SYPC	СМОЕВ	Y16560
Description						

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
Protein name	604	5826	148	Locus	3 Name	Acc#
Description NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
29353952 <u></u> c1 <u></u> 304	. 605	5827	592	1779	891	3.4e-89
Protein name				Locus	Name	Acc#
				gp:AF02	25396	AF025396
Description						
Vibrio anguillarum	rfb regio	n, partia	I sequenc	ce.	*	
		0	NT	ΔΔ		
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 3.0.5.22.9.6f2122		<u>AAID</u> 5828	Length		Score	Probability 5.9e-07
			Length	Length		
30522966 <u></u> £2 <u></u> 122			Length	Length	132 Name	5,9e-07
30522966 <u></u> £2 <u></u> 122			Length	Length 789 Locus	132 Name	5.9e-07 <u>Acc#</u>
30522966f2122 Protein name	. [606		Length	Length 789 Locus	132 Name	5.9e-07 <u>Acc#</u>
Protein name  Description	. [606		Length  262	Length 789  Locus	132 Name	5.9e-07 <u>Acc#</u>
Protein name  Description	. [606		Length	Length 789 Locus	132 Name	5.9e-07 <u>Acc#</u>
20.522966t2122  Protein name  Description  REGULATORY PROTEIN	TENI	5828	Length  262  NT Length	Length 789  Locus Sp:TEN	132 S Name L BACSU	5.9e-07 <u>Acc#</u> P25053
20.522966t2122  Protein name  Description  REGULATORY PROTEIN  ORF Name	TENI  NTID	5828 AAID	Length  262  NT Length	Length Locus Sp:TEN  AA Length	132 S Name L_BACSU Score	5.9e-07 Acc# P25053  Probability
Description  REGULATORY PROTEIN  ORF Name  32165905_f1_5	TENI  NTID  [607	5828 AAID	Length  262  NT Length	Length Locus Sp:TEN  AA Length	132 S Name Score  82 Name	5.9e-07 Acc# P25053  Probability  0.0018

		NT AA .	
ORF Name	NTID AAID	Length Length Score Pro	bability
33289500_c2_393	608 5830	307 924 1394 1	7e-142
Protein name		Locus Name	Acc#
glucose-1-phospha	te thymidyl transfer	gp:AF048749	AF048749
Description			
Bacteroides fragi sequence.	lis capsular polysa	ccharide biosynthesis operon,	complete
		NT AA	
ORF Name	NTID AAID	Length Length Score Pro	bability
34064010_f3_232	609 5831	478 1437 707 1	.1e-69
Protein name		Locus Name	Acc#
RNA methyltransfe	rase homolog yefA	pir:E69793	E69793
Description	-		
×			
ORF Name	NTID AAID	NT AA Length Length Score Pro	bability
34079635 <u></u> £1 <u>9</u>	610 5832	168 507	
Protein name		Locus Name	Acc#
Description			
NO-HIT	-		
	1	NT AA Score Pro	bability
ORF Name	NTID AAID	Length Length	
341735563482	611 5833	314 945 223 4	.7e-18
Protein name		Locus Name	Acc#
ADP-L-glycero-D-ma	enno-hentose-6-enime	erase pir:G70330	G70330
	amo nepeobe o epime	F	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
34245941 c3 490	612	5834	86   2	261		
Protein name			L	Locus	Name	Acc#
Description						
NO-HIT						
			TM	AA	Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	Length	<u>50010</u>	11000011101
3.43.8.46.3.1 <u></u> ±3 <u></u> .216	613	5835	205	518	319	1.4e-28
Protein name				Locus	Name	Acc#
				sp:THI	e_syny3	P72965
Description						•
PYROPHOSPHORYLASE	) (TMP-PPAS	E) (THIAM	INE-PHOSI	PHATE SYN	THASE)	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
35214385 <u></u> c2397	614	5836	297	394	182	6.8e-13
Protein name				Locus	Name	Acc#
glucosyl transfera	ise			gp:SMU5	2844	U52844
Description				<u> </u>	eTe	
Serratia marcescer putativeglycosyltra (waaQ),3-deoxy-mann glucosyltransferase Fpg(fpg) gene, part	ansferase, no-octuloso e (waaE), a	putative lonic acid	neptosyll transfera	III trans ase (waa <i>A</i>	۸) ,	s; and
			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length		
35348458 <u></u> c1 <u></u> 327	615	5837	119	860	93	0,00025
Protein name				Locus	Name	Acc#
unknown		-		gp:AF00	7381	AF007381
Description				l h		, v
Flavobacterium joh cds; and unknown ge		iding mot	ility pro	otein (gl	.dA) gene	e,complete

ORF Name	NTID AAID	NT AA Score Probability Length Length
36330078_c3_518	5838	96 291
Protein name		Locus Name Acc#
Description		
по-ніт		
ORF Name	NTID AAID	NT AA Score Probability Length Length
36601442±3215	617 5839	118 357 118 2.8e-07
Protein name		Locus Name Acc#
		gp:STYSTMF1 AF170176
Description		
Salmonella typhim	urium fragment STMF	1.
		277
ORF Name	NTID AAID	NT AA Score Probability Length Length
3.93.9512 <u></u> c1326	618 5840	909 2730 1830 4.0e-281
Protein name	-	Locus Name Acc#
Protein name	-	Locus Name Acc#  sp:PODK_CLOSY P22983
Protein name  Description		
,		
Description		Sp:PODK_CLOSY P22983
Description	NTID AAID	
Description DIKINASE)		Sp:PODK_CLOSY P22983  NT AA Score Probability
Description DIKINASE) ORF Name		Sp:PODK_CLOSY P22983  NT AA Score Probability Length Length
Description  DIKINASE)  ORF Name  3.9408.77t296	[619 [5841	Sp:PODK_CLOSY P22983  NT AA Score Probability Length Length 591 9.3e-60
Description  DIKINASE)  ORF Name  3.940877_t2_96  Protein name  precorrin-3 methy  Description	619 5841 lase	Sp:PODK_CLOSY P22983  NT AA Score Probability Length Length 591 9.3e-60  Locus Name Acc#

			NT	AA T	Score	Probability
ORF Name	NTID	AAID	Length	Length		
39642_c3_480	620	5842	196	91	841	6.7e-84
Protein name				Locus	Name	Acc#
dTDP-6-deoxy-D-gluc	cose-3,5 ep	imerase		gp:AF04	8749	AF048749
Description						
Bacteroides fragili sequence.	s capsular	polysaco	charide b	piosynthe	sis oper	con, complete
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
40930_f2_144	621	5843	140 4	23	391	3.2e-36
Protein name				Locus	Name	Acc#
conserved hypotheti	.cal protei	ı		pir:C75	256	C75256
Description						
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
ORF Name 4196081G2391		<u>AAID</u> 5844	Length		Score	Probability
			Length	Length 71	Score Name	Probability  Acc#
4196081 <u></u> c2 <u></u> 391			Length	Length 71		· · · · · · · · · · · · · · · · · · ·
4196081c2391 Protein name			Length	Length 71		· · · · · · · · · · · · · · · · · · ·
Protein name  Description			Length	Length 71	Name	Acc#
Protein name  Description			Length	Length 71 Locus		· · · · · · · · · · · · · · · · · · ·
Protein name  Description  NO-HIT	NTID	5844	Length  156  A  NT  Length	Length  TI  Locus	Name	Acc#
Protein name  Description  NO-HIT  ORF Name	NTID	AAID	Length  156  AT  Length	Length Locus  AA Length 311	Name Score	Acc# Probability
#196081_c2_391 Protein name Description NO-HIT ORF Name #334455_c3_463	NTID	AAID	Length  156  AT  Length	Length Locus  AA Length 311	Name Score 227 Name	Acc#  Probability  3.5e-19
#196081_c2_391 Protein name Description NO-HIT ORF Name #334455_c3_463	NTID	AAID	Length  156  AT  Length	Length Locus  AA Length 311 Locus	Name Score 227 Name	Acc#  Probability  3.5e-19  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4688828_c3_507	624	5846	96	291	87	0.0012	
Protein name				Locus	Name	Acc#	
unknown				gp:AF00	7381	AF007381	
Description							
Flavobacterium jocds; and unknown go		ding mot:	ility pro	otein (gl	.dA) gene	e,complete	*
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	-
4881512_f1_72	625	5847	153	162			
Protein name				Locus	3 Name	Acc#	
Description						-	
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	- 3.5
ORF Name 4881557c2404		<u>AAID</u> 5848	Length		Score	Probability	
		0	Length	Length		*	
4881557 <u></u> c2 <u></u> 404	protein sll	5848	Length	Length	105 S Name	0.016	i i
4881557c2404 Protein name  Na+/H+-exchanging	protein sll	5848	Length	Length Locus	105 S Name	0.016 <u>Acc#</u>	· · · · · · · · · · · · · · · · · · ·
Protein name  Na+/H+-exchanging antiporter:Na+/H+	protein sll	5848	Length  395	Length Locus	105 S Name 1414	0.016 <u>Acc#</u> S74414	2
Protein name  Na+/H+-exchanging antiporter:Na+/H+	protein sll	5848	Length	Length Locus	105 S Name	0.016 <u>Acc#</u>	2
4881557c2404 Protein name  Na+/H+-exchanging antiporter:Na+/H+  Description	protein sll	5848 -0689:Na+/	Length  395  /H+	Length  Locus  pir:S7	105 S Name 1414	0.016 <u>Acc#</u> S74414	
4881557c2404  Protein name  Na+/H+-exchanging antiporter:Na+/H+  Description  ORF Name	protein sll	5848 .0689:Na+,	Length  395  /H+  NT  Length	Length  Locus  pir:S7	105 S Name 4414 Score	0.016 Acc# S74414 Probability	, j
Protein name  Na+/H+-exchanging antiporter:Na+/H+  Description  ORF Name  4882755_t3_268	protein sliantiporter  NTID	5848 .0689:Na+,	Length  395  /H+  NT  Length	Length  Locus  pir:S7	105 S Name 4414 Score 103 S Name	0.016 Acc# S74414  Probability  0.0030	
Protein name  Na+/H+-exchanging antiporter:Na+/H+  Description  ORF Name  4882755_t3_268	protein sllantiporter  NTID  627  protein	AAID 5849	Length  395  /H+  NT Length  311	Length Locus  AA Length  Sacrate  Locus  gp:ZEF6	105 S Name 4414 Score 103 S Name GAP	0.016 Acc# S74414  Probability  0.0030 Acc#	

		$\frac{\text{NT}}{}$ $\frac{\text{AA}}{}$ Score	Probability
ORF Name	NTID AAID	Length Length	
4884635_c3_487	628 5850	266 801 403	1.7e-37
Protein name		Locus Name	Acc#
unknown		gp:AF144879	AF144879
Description			
Leptospira interr	ogans rfb locus, com	plete sequence.	
		7.7	<del></del>
ORF Name	NTID AAID	NT AA Score	Probability
4884712_c2_401	629 5851	254 765 636	3.5e-62
Protein name	·	Locus Name	Acc#
exodeoxyribonucle	ase	pir:B69126	B69126
Description			in Va
7			
ORF Name	NTID AAID	NT AA Score	Probability
4957512 c3 506		193 582	
		Locus Name	Acc#
Protein name		Hoods Hamo	
Description			
NO-HIT			
		NT AA C	Decelorly
ORF Name	NTID AAID	Length Length Score	Probability
5.04.757 <u></u> c2 <u></u> 3.73	631 5853	956 2871 383	3.0e-34
Protein name		Locus Name	Acc#
RcsC	THE RESERVE OF THE RE	gp:AF071215	AF071215
Description			
Proteus mirabilis	s regulator of swarm	ing behavior precursor ( (rcsC) gene, partialcds.	rsbA) and RcsB

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
5110325_c3_521	632	5854	381	.146	1054	1.8e-106
Protein name				Locus	s Name	Acc#
carboxynorspermi sll0873:protein s		ylase:pro	otein	pir:S7	7268	S77268
Description				.,,		
			NT	AA	Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	Length	9 ;	
5112802 <u></u> f3 <u></u> 207	633	5855	644	.935	423	2.8e-39
Protein name				Locus	s Name	Acc#
CbiD protein				gp:BMA	758	AJ000758
Description						
	adam 1 Clab gone				no won	
Bacillus megate:	rium ieko gene	omic sedue	ence, cor	Dalamili C	pperon.	
Bacillus megate:	rium 16kb gen	omic seque		AA	*	
ORF Name	NTID	AAID	NT Length		Score	Probability
- :	NTID		NT Length	AA	*	Probability
ORF Name	NTID	AAID	NT Length	AA Length	*	Probability  Acc#
ORF Name 6.0256.75£21.62	NTID	AAID	NT Length	AA Length	Score	
ORF Name 6.0256.75f2162  Protein name	NTID	AAID	NT Length	AA Length	Score	
ORF Name 6.0256.75_f2_162  Protein name Description	NTID	AAID	NT Length	AA Length	Score S Name	Acc#
ORF Name 6.0256.75_f2_162  Protein name Description	NTID	AAID	NT Length 75 2	AA Length 228 Locus	Score	
ORF Name 6.0256.75f2162  Protein name  Description  NO-HIT	NTID 634 NTID	<u>AAID</u> [5856	NT Length [75]	AA Length 228 Locus	Score S Name	Acc#
ORF Name 6.0256.75f2162 Protein name Description NO-HIT ORF Name	NTID 634 NTID	<u>AAID</u> [5856	NT Length [75]	AA Length  Locus  AA Length  1353	Score Score	Acc# Probability
ORF Name 6.0256.75f2162  Protein name Description NO-HIT  ORF Name 6.9138.75c1305	NTID	<u>AAID</u> [5856	NT Length [75]	AA Length  Locus  AA Length  1353	Score  Score  Score  602  Name	Acc# Probability 1.4e-58

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
7087642_c1_306	636	5858	61 ]	.86	. Namo	7 cc#
Protein name				Locus	: Name	Acc#
Description NO-HIT						
110 1111						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
910262 <u></u> f3 <u></u> 235	637	5859	305	918	670	8.8e-66
Protein name					s Name	Acc#
				sp:AMP	L_SYNY3	P53579
Description						
M)						1
		*	NT	AΑ		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 969165_c1_297	-	<u>AAID</u> 5860	Length		Score	1.1e-11
	-		Length	Length		
969165 <u></u> c1297	638	5860	Length	Length	182 s Name	1.1e-11
969165 <u></u> c1 <u></u> 297 Protein name	638	5860	Length	Length  1011  Locu	182 s Name	1.1e-11 <u>Acc#</u>
Protein name  conserved hypothe	638	5860	Length  336	Length  Locu  pir:F6	182 S Name	1.1e-11 <u>Acc#</u> F69035
Protein name  conserved hypothe	638	5860	Length	Length  1011  Locu	182 s Name	1.1e-11 <u>Acc#</u>
Protein name  conserved hypothe  Description	638 etical protei	5860 n MTH126	Length  336  NT Length	Length  Locu  pir:F6  AA  Length	182 s Name 9035 Score	1.1e-11 Acc# F69035  Probability  0.014
Protein name  conserved hypothe  Description  ORF Name	638 etical protei	2860 n MTH126	Length  336  NT Length	Length  Locu  pir:F6  AA  Length	182 S Name 9035 Score	Acc# F69035  Probability
Protein name  Conserved hypothe  Description  ORF Name	638 etical protei	2860 n MTH126	Length  336  NT Length	Length  Locu  pir:F6  AA  Length  849  Locu	182 s Name 9035 Score	1.1e-11 Acc# F69035  Probability  0.014
Protein name  Conserved hypothe  Description  ORF Name	NTID	AAID	Length  336  NT Length  282	Length Locu  pir:F6  AA Length Locu  sp:YBJ	Name 9035 Score 105 s Name	1.1e-11   Acc#   F69035   Probability   0.014   Acc#

Protein name	ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
Nypothetical protein PH1670   Description	11757880_c2_81	640	5862	146 4	41	108	3.2e-06
Description	Protein name				Locus	Name	Acc#
ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         641         5863         363         1092           Protein name         Locus Name         Acc#           Description         NO-HIT         AAID         Length         Length         Score         Probability           Protein name         NTID         AAID         Length         Length         Acc#           Protein name         Locus Name         Acc#           Description         NO-HIT         Length         Length         Score         Probability           NO-HIT         AAID         Length         Length         Score         Probability           LSE/B277_G3_S5	hypothetical protei	n PH1670			pir:F71	1047	F71047
ORF Name         NTID         AAID         Length         Length         Length         Formal Length           Protein name         Locus Name         Acc#           Description         NOTH         Locus Name         Acc#           NOTHIT         NOTH         AA         Length         Length         Score         Probability           IA12BA13_F1_8         642         5864         95         288         Protein name         Acc#           Description         NOTHIT         NOTHIT         AAID         Length         Length         Score         Probability           NRF Name         NTID         AAID         Length         Length         I.1e-12           Protein name         Locus Name         Acc#           Serine-rich protein         pir:T39903         T39903           Description           ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         NTID         AAID         Length         Length         Score         Probability           Protein name         Length         Length         Length         Score         Probability           Locus Name         Acc#         S	Description						
ORF Name         NTID         AAID         Length         Length         Length         Formal Length           Protein name         Locus Name         Acc#           Description         NOTH         Locus Name         Acc#           NOTHIT         NOTH         AA         Length         Length         Score         Probability           IA12BA13_F1_8         642         5864         95         288         Protein name         Acc#           Description         NOTHIT         NOTHIT         AAID         Length         Length         Score         Probability           NRF Name         NTID         AAID         Length         Length         I.1e-12           Protein name         Locus Name         Acc#           Serine-rich protein         pir:T39903         T39903           Description           ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         NTID         AAID         Length         Length         Score         Probability           Protein name         Length         Length         Length         Score         Probability           Locus Name         Acc#         S				NT	AA		Desch ab 11 i has
Protein name   Locus Name   Acc#	ORF Name	NTID	AAID			score	Probability
Description   NO-HIT	1446.0.957t14	641	5863	363	L092		
NO-HIT	Protein name				Locus	s Name	Acc#
ORF Name         NTID         AAID         Length         Length         Score         Probability           1A7.2BA13_£1_8         642         5864         95         288         Locus Name         Acc#           Protein name         NT DESCRIPTION         NT DESCRIPTION         AAID DESCRIPTION         NTID DESCRIPTION         AAID DESCRIPTION         Length DESCRIPTION         Length DESCRIPTION         NTID DESCRIPTION         AAID DESCRIPTION         DESCRIPTION         DESCRIPTION         NTID DESCRIPTION         AAID DESCRIPTION         Length DESCRIPTION         Score Probability           ORF Name         NTID DESCRIPTION         AAID DESCRIPTION         Length DESCRIPTION         Score Probability           Protein name         NTID DESCRIPTION         AAID DESCRIPTION         Length DESCRIPTION         Score Probability           DESCRIPTION         AAID DESCRIPTION         AAID DESCRIPTION         AAID DESCRIPTION         AAID DESCRIPTION	Description						
NTID   AAID   Length   Lengt	NO-HIT						
NTID   AAID   Length   Lengt				NTITI	7.7		
Protein name         Locus Name         Acc#           Description         NO-HIT         AA         Score         Probability           ORF Name         NTID         AAID         Length         Length         Score         Probability           1587.87.77_c3_95         643         5865         232         699         178         1.1e-12           Protein name         Locus Name         Acc#           serine-rich protein         pir:T39903         T39903           Description         NTID         AAID         Length         Length         Score         Probability           20.741.053_f2_25         644         5866         240         723         1219         5.9e-124           Protein name         Locus Name         Acc#           BatC         GP:AF116251         AF116251	ORF Name	NTID	AAID			Score	Probability
Description	14728413_f1_8	642	5864	95	288		
NO-HIT         NT Length Length Length Length         Score Length Length         Probability           158787777c395	Protein name				Locu	s Name	Acc#
NO-HIT         NT Length Length Length Length         Score Length Length         Probability           158787777c395	Description						
ORF Name         NTID         AAID         Length         Length         Score         Probability           158.78.77.7395							
ORF Name         NTID         AAID         Length         Length         Score         Probability           158.78.77.7395	1						
Protein name	ORF Name	NTID	AAID			Score	Probability
Serine-rich protein	15878777c395	. 643	5865	232	699	178	1.1e-12
Description         NT         AA         Score         Probability           ORF Name         NTID         AAID         Length         Length         Score         Probability           20741053_f2_25         644         5866         240         723         1219         5.9e-124           Protein name         Locus Name         Acc#           BatC         gp:AF116251         AF116251	Protein name	J L		<u></u>	Locu	s Name	Acc#
ORF Name         NTID         AAID         Length         Length         Score         Probability           20741053_f2_25         644         5866         240         723         1219         5.9e-124           Protein name         Locus Name         Acc#           BatC         gp:AF116251         AF116251           Description	serine-rich protei	n			pir:T3	9903	T39903
ORF Name         NTID         AAID         Length         Length         Score         Probability           20.7.410.53f225	Description						-
ORF Name         NTID         AAID         Length         Length         Score         Probability           20.7.410.53f225						· · · · · · · · · · · · · · · · · · ·	
20.7410.53f225	ORF Name	NTID	AAID			Score	Probability
Protein name         Locus Name         Acc#           BatC         gp:AF116251         AF116251           Description         Description         AF116251		644	5866	240	723	1219	5.9e-124
BatC gp:AF116251 AF116251  Description		<u> </u>			Locu	s Name	Acc#
				<u></u>	gp:AF1	16251	AF116251
	L					*	
		is batl or	peron, com	plete se	quence.		:

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
209688_£3_38	645	5867	63 1	.92			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
22027 <u>_</u> £3 <u>_</u> 31	646	5868	355 1	.068	683	3.7e-67	]
Protein name				Locus	Name	Acc#	
				sp:GCP_	HAEIN	P43764	
Description							
(GLYCOPROTEASE)			*		-		
			NT	AA	~	D 1 1 1 - 1 - 1	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
2203.96.92 <u></u> .£2 <u></u> 24	647	5869	216	51	1014	3.1e-102	
Protein name				Locus	s Name	Acc#	
BatB	ı X		-	gp:AF1	16251	AF116251	
Description							
Bacteroides fragil	is batl ope	eron, comp	olete sec	quence.			
Bacteroides fragil	is batI ope	eron, comp					
Bacteroides fragil ORF Name	is batI ope	eron, comp	NT Length	AA Length	Score	Probability	
	NTID		<u>NT</u> Length	AA	Score	Probability 3.6e-76	
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length			
ORF Name 22297140t13	NTID	AAID	<u>NT</u> Length	AA Length 984	768	3.6e-76	
ORF Name 22297140t13	NTID	AAID	<u>NT</u> Length	AA Length 984	768 s Name	3.6e-76	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
23649052_f1_9	649	5871	399	200	166	1.2e-11
Protein name				Locus	s Name	Acc#
				sp:Y53	L_METJA	Q57951
Description						
HYPOTHETICAL PROTE	IN MJ0531					
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
23834376_c2_84	650	5872	122	369	122	1.0e-07
Protein name				Locu	s Name	Acc#
hypothetical prote	ein APE1982			pir:H7	2500	H72500
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 24259700f16		<u>AAID</u> 5873	Length		<u>Score</u>	Probability
			Length	Length		
24259700 <u></u> £1 <u></u> .6			Length	Length	3076 s Name	0.0
Protein name  BatD  Description	651	5873	Length	Length  1848  Locu  gp:AF1	3076 s Name	0.0 <u>Acc#</u>
24259700f16 Protein name  BatD	651	5873	Length	Length  1848  Locu  gp:AF1	3076 s Name	0.0 <u>Acc#</u>
Protein name  BatD  Description	651	5873	Length 615	Length  Locu  Locu  gp:AF1  quence.	3076 s Name	0.0 <u>Acc#</u>
Protein name  BatD  Description  Bacteroides fragi	651  lis batl op	peron, com	Length 615  plete se  NT  Length	Length  Locu  Gp:AF1  quence.  AA  Length	3076 s Name 16251 Score	Acc# AF116251  Probability
Protein name  BatD  Description  Bacteroides fragi  ORF Name	651  lis batl op	5873	Length 615  plete se  NT  Length	Length  Locu  gp:AF1  quence.  AA  Length	3076 s Name 16251 Score	Drobability  1.5e-13
Protein name  BatD  Description  Bacteroides fragi  ORF Name  2440.7537f221	651 lis batl op  NTID 652	peron, com	Length 615  plete se  NT  Length	Length  Locu  gp:AF1  quence.  AA  Length  282  Locu	3076 s Name 16251 Score 177 s Name	Acc# AF116251  Probability  1.5e-13 Acc#
Protein name  BatD  Description  Bacteroides fragi  ORF Name	651 lis batl op  NTID 652	peron, com	Length 615  plete se  NT  Length	Length  Locu  gp:AF1  quence.  AA  Length	3076 s Name 16251 Score 177 s Name	Drobability  1.5e-13

			$\underline{\mathtt{NT}}$	<u>AA</u>	Score	Probability	
ORF Name	NTID	AAID	<u>Length</u>	Length			
24415903_f1_7	653	5875	279	340	1381	4.0e-141	
Protein name				Locus	Name	Acc#	
BatE				gp:AF11	16251	AF116251	
Description							
Bacteroides fragili	s batI ope	eron, com	plete sec	quence.			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24823562_f1_2	654	5876	68 2	207	157	2.0e-11	]
Protein name				Locus	s Name	Acc#	
-				sp:RK3	3_ODOSI	P49565	
Description					*		
CHLOROPLAST 50S RI	BOSOMAL PR	OTEIN L33			ő		
CHLOROPLAST 50S RI	BOSOMAL PR	OTEIN L33			ñ		
	BOSOMAL PR	AAID	NT Length	AA Length	Score	Probability	
ORF Name  25975010c163			<u>NT</u> Length		Score	Probability	•
ORF Name	NTID	AAID	<u>NT</u> Length	Length	Score s Name	Probability  Acc#	
ORF Name 25975010_c1_63  Protein name	NTID	AAID	<u>NT</u> Length	Length	·.		
ORF Name 25975010c1_63	NTID	AAID	<u>NT</u> Length	Length	·.		*
ORF Name 25975010_c1_63  Protein name	NTID	AAID	<u>NT</u> Length	Length	·.		*
ORF Name  25975010_c1_63  Protein name  Description	NTID	AAID	NT Length	Length 3390 Locus	s Name	Acc#	*
ORF Name  25975010_c1_63  Protein name  Description	NTID	AAID	<u>NT</u> Length	Length	·.		*
ORF Name  25975010_c1_63  Protein name  Description  NO-HIT	NTID 655 NTID	<u>AAID</u> 5877	NT Length 1129	Length  Locus	s Name	Acc#	
ORF Name  25975010c163  Protein name  Description  NO-HIT  ORF Name	NTID 655 NTID	AAID  AAID	NT Length 1129	Length  Locus  AA  Length  Langth	s Name	Acc# Probability	*
ORF Name  25975010_c1_63.  Protein name  Description  NO-HIT  ORF Name  272212_F3_33.	NTID 655 NTID	AAID  AAID	NT Length 1129	Length  Locus  AA  Length  Langth	Score  201  S Name	Acc# Probability 1.6e-12	

		NT	AA Canasa	Probability
ORF Name	NTID AAII	Length Le	ength Score	PIODADITICY
3149069_c1_40	657 5879	847 254	4304	0.0
Protein name			Locus Name	Acc#
DNA gyrase A subu	nit	a	p:AB017712	AB017712
Description				
Bacteroides fragi	lis gyrA gene for	DNA gyrase A	subunit, comp.	letecds.
		NT	AA	_ 1 1 1 1 1 1
ORF Name	NTID AAII		ength Score	Probability
33397127_f1_5	658 5880	331 996	1656	2.9e-170
Protein name			Locus Name	Acc#
BatA		a	p:AF116251	AF116251
Description				
Bacteroides fragi	lis batl operon, o	complete seque	nce.	
Bacteroides fragi	lis batl operon, o			
Bacteroides frag	NTID AAII	NT	AA Score	Probability
	NTID AAII	NT	AA Score	Probability 6.7e-36
ORF Name	NTID AAII	NT D Length Le	AA Score	
ORF Name 34101702_f2_23	NTID AAII	NT D Length Le	AA Score	6.7e-36
ORF Name 34101702_f2_23	<u>NTID</u> <u>AAII</u>	NT D Length Le	AA Score  388  Locus Name	6.7e-36 <u>Acc#</u>
ORF Name  34101702_t2_23  Protein name  conserved hypothe	<u>NTID</u> <u>AAII</u>	NT D Length Le	AA Score  388  Locus Name	6.7e-36 <u>Acc#</u>
ORF Name  34101702_t2_23  Protein name  conserved hypothe	<u>NTID</u> <u>AAII</u>	NT Length Le 289 870	AA Score  388  Locus Name	6.7e-36 <u>Acc#</u>
ORF Name  341017.02_f2_23  Protein name  conserved hypothe  Description	NTID AAII 659 5881  Stical protein BBO: NTID AAI	NT Length Le 289 870	AA Score  388  Locus Name  1r:G70121  AA Score	6.7e-36 Acc# G70121
ORF Name  341017.02_f2_23  Protein name  conserved hypothe  Description  ORF Name	NTID AAII 659 5881  Stical protein BBO: NTID AAI	NT D Length Le 289 870 175 F  NT D Length Le	AA Score  388  Locus Name  1r:G70121  AA Score	6.7e-36  Acc#  G70121  Probability
ORF Name  34101702_f2_23  Protein name  conserved hypothe Description  ORF Name  34266968_f3_32  Protein name	NTID AAII 659 5881  Stical protein BBO: NTID AAI	NT Length	AA Score  Locus Name  ir:G70121  AA Score  942	6.7e-36 Acc# G70121  Probability  1.3e-94

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
34564376_f1_1	661	5883	415	.248	603	1.1e-58
Protein name				Locus	s Name	Acc#
hypothetical prote	in			pir:S76	5561	S76561
Description		*				
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
3912925_f3_34	662	5884	334	L005	821	8.8e-82
Protein name				Locus	s Name	Acc#
probable moxR prot	ein			pir:B70	0874	B70874
Description				200	-	
			NT	AA		_ , , , , , , , ,
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 4961537G387	7	<u>AAID</u> 5885	Length		Score	Probability 6.9e-09
	7		Length	Length		
4961537c387	663	5885	Length 418	Length	164 s Name	6.9e-09
4961537c387 Protein name	663	5885	Length 418	Length 1257 Locu	164 s Name	6.9e-09 <u>Acc#</u>
Protein name  conserved hypothet	663	5885	Length 418	Length 1257 Locu	164 s Name	6.9e-09 <u>Acc#</u>
Protein name    Conserved hypothet   Description	MTID	5885 in aq_854	Length  418  NT Length	Length Locu. pir:B7	164 s Name 0374	6.9e-09 <u>Acc#</u> B70374
Protein name    Conserved hypothet	MTID	5885 in aq_854 	Length  418  NT Length	Length Locu. pir:B7  AA Length	164 s Name 0374 Score	Acc# B70374  Probability
Protein name  Conserved hypothet  Description  ORF Name	MTID	5885 in aq_854 	Length  418  NT Length	Length Locu  Dir:B7  AA Length  Locu  Locu  Locu	164 s Name 0374 Score	6.9e-09 Acc# B70374  Probability 4.2e-11
Protein name  Conserved hypothet  Description  ORF Name	MTID	5885 in aq_854 	Length  418  NT Length	Length Locu  Dir:B7  AA Length  Locu  Locu  Locu	164 S Name 0374 Score 154 S Name	6.9e-09  Acc#  B70374  Probability  4.2e-11  Acc#

•			NT	AA		B
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
7072675_f3_36	665	5887	146 4	41	687	1.4e-67
Protein name				Locus	Name	Acc#
BatB				gp:AF11	6251	AF116251
Description						
Bacteroides fragi	lis batI op∈	eron, com	plete seq	uence.		
		7.	NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>	BCOLE	riobability
10562517_f3_78	666	5888	71 2	16		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
12129682 <u></u> £2 <u></u> 67	667	5889	516 1	.551	319	2.6e-26
Protein name		*		Locus	s Name	Acc#
lipase-like prote	in ,			pir:A6	4706	A64706
Description					-	
	*					
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
12679062 <u>t3</u> 97	668	5890	66 2	201		
Protein name			-	Locu	s Name	Acc#
Description						*
NO-HIT		-				
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
13834500 <u></u> 196	669	5891	144	135	203	2.7e-16
Protein name		<u> </u>	<u> </u>	Locu	s Name	Acc#
hypothetical prot	ein BB0530			pir:A7	0166	A70166
Description	· ·			J L		

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
17086686_±3_95	670 5	892	268	307		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
20585963 <u>f1</u> 19	671 5	893	413 1	.242	226	1.8e-16
Protein name				Locus	Name	Acc#
hypothetical prote	in jhp1380			pir:G71	815	G71815
Description						
			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length	<u>50010</u>	22000022201
2151552 <u></u> c2 <u></u> 146	672 5	894	250	753	104	2.7e-05
Protein name				Locus	Name	Acc#
cytochrome b				gp:AF01	7516	AF017516
Description						
Bombus pascuorum o mitochondrial prote			gene, mit	ochondri	.al genee	encoding
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
22473516 <u></u> c3 <u></u> 186	673 5	895	500	503	1054	1.8e-106
Protein name				Locus	Name	Acc#
				sp:CBI	SALTY	Q05597
Description						
COBYRIC ACID SYNT	HASE		**			

	NTID AAID	NT AA Score Probability Length Length
ORF Name		
24269180_f2_71	674 5896	
Protein name		Locus Name Acc#
hypothetical prot	ein jhp1379	pir:F71815 F71815
Description		
		NT AA Ggoro Probability
ORF Name	NTID AAID	Length Length Score Probability
24317806t253	675 5897	321 966 436 5.5e-41
Protein name		Locus Name Acc#
nicotinate-nucleo	otidedimethylbenz nsferase	imidazole pir:A75577 A75577
Description		
		NT AA Grama Probability
ORF Name	NTID AAID	Score Probability
ORF Name 24345167£14		Score Probability
*		Length Length Flobability
2434516.7£14  Protein name  Cobinamide kinas	676 5898 e / cobinamide phos	Length Length Score Flobability  206 621 310 1.2e-27  Locus Name Acc#
2434516.7t14 Protein name cobinamide kinas guanylyltransfera	676 5898 e / cobinamide phos	Length Length Score Flobability  206 621 310 1.2e-27  Locus Name Acc#
2434516.7£14  Protein name  Cobinamide kinas	676 5898 e / cobinamide phos	Length Length Score Flobability  206 621 310 1.2e-27  Locus Name Acc#
2434516.7t14 Protein name cobinamide kinas guanylyltransfera	676 5898 e / cobinamide phos	Length Length  206 621 310 1.2e-27  Locus Name Acc#  phate pir:S52220  NT AA Score Probability
2434516.7t14 Protein name cobinamide kinas guanylyltransfera	676 5898 e / cobinamide phos	Length Length    206   621   310   1.2e-27     Locus Name   Acc#     phate   pir:S52220     NT   AA   Score   Probability     Length Length   Lengt
2434516.7t14	676 5898  e / cobinamide phos	Length         Length         Score         Flobability           206         621         310         1.2e-27           Locus Name         Acc#           phate         pir:S52220           NT Length         AA Length         Score         Probability           502         1509         1238         5.7e-126
Protein name  cobinamide kinas guanylyltransfera  Description  ORF Name  24417212_c3_197	676 5898  e / cobinamide phospise  NTID AAID 677 5899	Length         Length         Score         Flobability           206         621         310         1.2e-27           Locus Name         Acc#           phate         pir:S52220           NT Length         Length         Score         Probability           Length         Length         5.7e-126           Locus Name         Acc#
Protein name  cobinamide kinas guanylyltransfera  Description  ORF Name  24417212_c3_197	676 5898  e / cobinamide phos se  NTID AAID 677 5899  gase, pros:proly1-t	Length         Length         Score         Flobability           206         621         310         1.2e-27           Locus Name         Acc#           phate         pir:S52220           NT Length         Length         Score         Probability           Length         Length         5.7e-126           Locus Name         Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24641903_c2_166  Protein name	678 59	900	168 5	07 Locus	Name	Acc#	
Description					0	-	
NO-HIT							]
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24666005 <u></u> c2 <u></u> 156	. 679 59	901	114 3	45			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			275	7.7			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24822213_f2_68	680 5	902	333 1	002	664	3.8e-65	ý.
Protein name				Locus	Name	Acc#	
immunoreactive 36	kDa antigen J	PG14	*	gp:AF14	:5798	AF145798	
Description							
Porphyromonas ging complete cds.	ivalis strai	n W50 im	munoreac	tive 36	kDa anti	genPG14 gene,	
			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
24822688t266	681 5	903	136 4	11	114	7.3e-07	
Protein name				Locus	Name	Acc#	
hypothetical prote	in			pir:S76	776	S76776	
Description							

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
				98	178	2.0e-13
25401437_c1_115	682	5904	103			
Protein name				Locus	: Name	Acc#
				sp:YJJ1	_HAEIN	P44520
Description						
HYPOTHETICAL PROTE	EIN HI0108					-
			NT	AA	G	Dechabilites
ORF Name	NTID	AAID	Length	Length	Score	Probability
30082887_c2_145	683	5905	279	340	238	5.3e-20
Protein name				Locus	Name	Acc#
				sp:YJJI	P_ECOLI	P39402
Description						2
HYPOTHETICAL 30.5	אדים אם איי	- דא האם אד	BGLJ TNT	RCENTC F	REGION ()	F277)
HIPOIDELICAL 30.3	TO THOTHER	TIV DIVIT	DOE0			
HYPOTHETICAL 30.3	TO TROTLET	11/ 21/11				
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
	NTID		<u>NT</u> Length	AA		
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 31657080c2144	NTID	AAID	<u>NT</u> Length	AA Length 1743	Score	Probability 5.5e-112
ORF Name 31657080c2144	NTID	AAID	<u>NT</u> Length	AA Length 1743	Score 1106 Name	Probability 5.5e-112
ORF Name 31657080_c2_144  Protein name	NTID	<u>AAID</u> 5906	NT Length 580	AA Length 1743 Locus	Score  1106  Name  ECOLI	Probability  5.5e-112  Acc#
ORF Name 31657080_c2_144  Protein name  Description	NTID	<u>AAID</u> 5906	NT Length 580	AA Length 1743 Locus sp:YID	Score  1106  Name  ECOLI	Probability  5.5e-112  Acc#
ORF Name 31657080_c2_144  Protein name  Description	NTID	<u>AAID</u> 5906	NT Length 580	AA Length 1743 Locus	Score  1106  Name  ECOLI	Probability  5.5e-112  Acc#
ORF Name 31657080_c2_144  Protein name  Description HYPOTHETICAL 58.9	NTID 684  KD PROTEIN NTID	AAID 5906	NT Length 580 [ IBPB INT]	AA Length 1743  Locus sp:YID1 ERGENIC H	Score  1106  Name E_ECOLI	Probability  5.5e-112  Acc#  ORFA)
ORF Name  31657080_c2_144  Protein name  Description  HYPOTHETICAL 58.9  ORF Name	NTID 684  KD PROTEIN NTID	AAID 5906 IN GLVC-	NT Length  580  IBPB INT  NT Length	AA Length  Locus  Sp:YIDI  ERGENIC I  AA Length	Score  1106  Name E_ECOLI  REGION (G	Probability  5.5e-112  Acc#  ORFA)  Probability
ORF Name  31657080_c2_144  Protein name  Description  HYPOTHETICAL 58.9  ORF Name  32228452_c3_172	NTID   684  KD PROTEIN NTID   685	AAID 5906 IN GLVC- AAID 5907	NT Length  580  IBPB INT  NT Length	AA Length  Locus  Sp:YIDI  ERGENIC I  AA Length	Score  [1106 S Name E_ECOLI  REGION (0  Score  [282 S Name	Probability  5.5e-112  Acc#  ORFA)  Probability  1.2e-24

		NT AA Garre	
ORF Name	NTID AAID	Length Length Score	Probability
33642211_f1_9	686 5908	260 783 156	3.7e-11
Protein name		Locus Name	Acc#
probable phosphog	lycerate mutase	pir:B75539	B75539
Description			
		7.7	
ORF Name	NTID AAID	NT AA Score Length Length	Probability
34195888 <u>c3</u> 190	687 5909	325 978 489	1.3e-46
Protein name		Locus Name	Acc#
		sp:COBD_PSEDE	P21634
Description			
COBD PROTEIN	No.		
		NT AA	
ORF Name	NTID AAID	Length Length Score	Probability
422125 <u></u> £1 <u></u> .16	688 5910	132 399 93	0.0029
Protein name		Locus Name	Acc#
beta-tropomyosin		pir:S23470	S23470
Description		. *	
		NT AA	
ORF Name	NTID AAID	NT AA Score Length Length	Probability
46.8.9.027 <u></u> c319.5	689 5911	1084 3255 895	1.3e-89
Protein name		Locus Name	Acc#
tricorn protease		gp:TAU72850	U72850
Description			
Thermoplasma acidophilum GTP-binding protein and tricorn protease(TRI) genes, complete cds.			

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4807062_c1_116	690	5912	448 1	.347	657	2.1e-64	
Protein name				Locus	Name	Acc#	
cobyrinic acid	a,c-diamide sy	nthase		pir:A7	619	A75619	
Description							
			NT	AA	Caoro	Probability	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
484451 <u></u> c2 <u></u> 149	691	5913	821 2	2466	436	1.3e-37	
Protein name					s Name	Acc#	
two component s	ensor			gp:AF0	30352	AF030352	
Description			345				
Pseudomonas aer	ruginosa two co	mponent	sensor ()	LemA) ger	ne, part	lalcds.	
ODE None	NULLO	7)	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name	NTID	AAID	Length	Length			7
5116586 <u></u> c1 <u></u> 131		<u>AAID</u> 5914	Length	Length	296	3.8e-26	
5116586_c1_131 Protein name			Length	Length  Tocus	296 s Name	3.8e-26 <u>Acc#</u>	
5116586_c1_131 Protein name			Length	Length	296 s Name	3.8e-26	
5116586_c1_131 Protein name	imurium alpha-	5914	Length  289  [289]  -5'-phosp	Length  Locus  Locus  phate pho	296 s Name 90625 ospatase	3.8e-26 Acc# U90625 CobC(cobC)	
Protein name  CobD  Description  Salmonella typh gene, partial co	nimurium alphads and putative	5914 ribazole aminotr	Length  289  -5'-phosp ansferase	Length  Locus  Gp:STU  Chate phose CobD (company)	296 s Name 90625 ospatase	3.8e-26 Acc# U90625 CobC(cobC)	
Protein name  CobD  Description  Salmonella typergene, partial cocces.  ORF Name	nimurium alphads and putative	ribazole e aminotr	Length  289  -5'-phosp ansferase  NT Length	Length  Locus  AA  Length	296  Name  90625  Dispatase cobD) gene	3.8e-26  Acc#  U90625  CobC(cobC)  e, complete	* *
Protein name  CobD  Description  Salmonella typergene, partial cocces.  ORF Name  5172508_f1_34	nimurium alphads and putative	5914 ribazole aminotr	Length  289  -5'-phosp ansferase  NT Length	Length  Locus  Gp:STU  Chate phose CobD (of AA Length  B12	296 s Name 90625 ospatase cobD) gene	3.8e-26  Acc#  U90625  CobC(cobC) e, complete  Probability	* *
Protein name  CobD  Description  Salmonella typh gene, partial cocds.  ORF Name  5172508_f1_34	nimurium alphads and putative	ribazole e aminotr	Length  289  -5'-phosp ansferase  NT Length	Length  Locus  Gp:STU  Chate phose CobD (of AA Length  B12	296  Name  90625  Dispatase cobD) gene	3.8e-26  Acc#  U90625  CobC(cobC)  e, complete	* *
Protein name  CobD  Description  Salmonella typergene, partial cocces.  ORF Name  5172508_f1_34	nimurium alphads and putative	ribazole e aminotr	Length  289  -5'-phosp ansferase  NT Length	Length  Locus  Gp:STU  Chate phose CobD (of AA Length  B12	296 s Name 90625 ospatase cobD) gene	3.8e-26  Acc#  U90625  CobC(cobC) e, complete  Probability	* *

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
5181263_f2_54	694	5916	250	53	268	3.5e-23
Protein name				Locus	Name	Acc#
cobalamin synthase				pir:H7	576	H75576
Description					×	
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
812510 <u></u> c2164	. 695	5917	66 2	201		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT						
				7.7		
ORF Name			NT	$\underline{AA}$	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>	50020	
10042126f3172	<u>NTID</u>	<u>AAID</u> 5918		Length	135	4.3e-09
				531		
10042126 <u></u> .‡3 <u></u> .172	696			531	135 Name	4.3e-09
10042126f3172 Protein name	696			Locus	135 Name	4.3e-09 Acc#
Protein name  hypothetical protes	. [696 in	5918	176	Locus	135 Name 18930	4.3e-09 Acc# Y18930
Protein name  hypothetical protein Description	. [696 in	5918	176 S	Locus gp:SSU	135 Name 18930	4.3e-09 Acc# Y18930
Protein name  hypothetical protein Description	. [696 in	5918	176	Locus	135 Name 18930	4.3e-09 Acc# Y18930
Protein name  hypothetical protein Description  Sulfolobus solfata	in ricus 281  NTID	5918 kb genomi	DNA fra  NT  Length	Locus gp:SSU	135 S Name 18930 Strain P2	4.3e-09 Acc# Y18930
Protein name  hypothetical protein  Description  Sulfolobus solfata:  ORF Name	in ricus 281  NTID	kb genomic	DNA fra  NT  Length	Locus  gp:SSU  agment, s  AA  Length	135 S Name 18930 Strain P2	4.3e-09 Acc# Y18930
Protein name  hypothetical protein  Description  Sulfolobus solfata:  ORF Name  11799076_c2_266	in ricus 281  NTID	kb genomic	DNA fra  NT  Length	Locus  gp:SSU  agment, s  AA  Length	135 S Name 18930 Strain P2 Score	4.3e-09  Acc#  Y18930  2.  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
1199075_c2_263	698	5920	64	.95		- "	
Protein name				Locus	Name	Acc#	
Description		_					
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
1359507 <u></u> c2 <u></u> 282	699	5921	434	1305	1073	1.7e-108	
Protein name				Locus	s Name	Acc#	
				sp:HIS	x_ECOLI		
Description						·.	
HISTIDINOL DEHYDRO	OGENASE, (H	DH)		0			
			NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
ORF Name 1370937c3383	-	<u>AAID</u> 5922	Length	Length	×		
	-		Length	Length	Score s Name	Probability  Acc#	
1370937 <u></u> c3 <u></u> 383	-		Length	Length	×		
1370937c3383 Protein name	-		Length	Length	×		
Protein name  Description	-		Length	Length	s Name	Acc#	
Protein name  Description	-		Length	Length 360 Locu	×		]
Protein name  Description  NO-HIT	[700 <u>NTID</u>	5922	Length  119  NT Length	Length 360 Locu	s Name	Acc#	]
Protein name  Description  NO-HIT  ORF Name	[700 <u>NTID</u>	5922 AAID	Length  119  NT Length	Length  Locu  AA  Length	s Name	Acc#	]
Protein name  Description  NO-HIT  ORF Name  13.756515f2119	[700 <u>NTID</u>	5922 AAID	Length  119  NT Length	Length  Locu  AA  Length	s Name	Acc# Probability	

L mil
24.45 Milli
gra
an dati and
N IN
2 H
f 25
ş
fi
tite first tites
tite first tites

		NT AA Score Probability
ORF Name	NTID AAID	Length Length
13786251_f1_58	702 5924	636 1911 386 1.8e-33
Protein name		Locus Name Acc#
histidine kinase		gp:AF114442 AF114442
Description		
Nostoc punctiform	e histidine kinase	(hepK) gene, complete cds.
		NT AA Grand Drobobility
ORF Name	NTID AAID	Length Length Score Probability
14630063_c3_356	703 5925	389 1170 880 4.9e-88
Protein name		Locus Name Acc#
		sp:HIS7_HAEIN P44327
Description		
	1	
ORF Name	NTID AAID	NT AA Score Probability Length Length
	704   5926	938 2817 2717 1.1e-282
Protein name		Locus Name Acc#
B12-dependent		gp:ECOUW89 U00006
Description		* ***
E. coli chromoson	nal region from 89.2	2 to 92.8 minutes.
		NICO 7.7)
ORF Name	NTID AAID	$rac{ ext{NT}}{ ext{Length}} rac{ ext{AA}}{ ext{Length}} rac{ ext{Score}}{ ext{Score}} rac{ ext{Probability}}{ ext{Probability}}$
15085902 <u></u> £2 <u></u> 125	705 5927	788 2367 1213 2.8e-129
Protein name	-	Locus Name Acc#
		sp:RHO_PSEFL P52155
Description		×
TRANSCRIPTION TE	RMINATION FACTOR RHO	

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length Scor	e <u>Probabilit</u>	У
 15629642_ <u>f</u> 2_79	706	5928	380 1	143 113	0.0067	
Protein name			<u></u>	Locus Name	Acc#	
				gp:PFMAL3P2		
Description						
Plasmodium falci	parum MAL3P2.	complete	e seguenc	e.		
Flasmodium rater	pazam rzim ,					
ORF Name	NTID	AAID	<u>NT</u> Length	AA Scor	e <u>Probabilit</u>	У
1995452_c3_340	707	5929	133 4	88	0.0023	
Protein name				Locus Name	Acc#	
				gp:SYCPURT	L36958	3
Description						
protein D2 (psbD)	, gene, s					
			NT	AA Scor	re Probabilit	Ϋ́
ORF Name	NTID	AAID	Length	Length Scor		
ORF Name 20119062_f3_159		<u>AAID</u> 5930	Length	Length Scor	1.1e-123	
20119062_f3_159 Protein name	708		Length	Length Scor	1.1e-123	3
20119062 <u></u> £3 <u></u> 159	708		Length	Length Scor	1.1e-123	3
20119062_f3_159 Protein name	708		Length	Length Scor	1.1e-123	3
20119062_f3_159  Protein name  hypothetical pro	708		Length	Length Scor	1.1e-123 Acc# D70680	0
20119062_t3_159  Protein name  hypothetical production	708  Stein Rv2438c  NTID	5930	Length 642  NT Length	Length Scor  1929 1069  Locus Name  pir:D70680	1.1e-123 Acc# D70680	0
20119062_f3_159  Protein name  hypothetical pro  Description  ORF Name	708  Stein Rv2438c  NTID	<u>AAID</u>	Length 642  NT Length	Length Scor  Locus Name  Locus Name  AA  Length Scor  AA  Length Scor	1.1e-123  Acc# D70686  Probabilit  0.0091	0 EY
20119062_f3_159  Protein name  hypothetical pro  Description  ORF Name  21520006_c2_301	708  Stein Rv2438c  NTID  709	<u>AAID</u>	Length 642  NT Length	Length Scor  Locus Name  Locus Name  AA  Length Scor  AA  Length 101	1.1e-123  Acc# D70686  Probabilit  0.0091  Acc#	0 EY
20119062_f3_159  Protein name  hypothetical product on Description  ORF Name  21520006_c2_301  Protein name	708  Stein Rv2438c  NTID  709	<u>AAID</u>	Length 642  NT Length	Length Scor  Locus Name  Locus Name  AA  Length Scor  AA  Length Scor  Locus Name	1.1e-123  Acc# D70686  Probabilit  0.0091  Acc#	0 EY

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
2161286_f3_160	710	5932	192 5	I.ogus	Name	Acc#	
Protein name  Description				<u>посив</u>	Trans	<u> </u>	
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
21641562_c2_271	711	5933	198	597			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 22289781_f3_195	<u>NTID</u>	<u>AAID</u> 5934	Length		<u>Score</u> 213	Probability 1.5e-13	_
			Length	Length 2031 Locus	213 S Name	1.5e-13 <u>Acc#</u>	
22289781 <u>f3</u> 195 Protein name			Length	Length 2031 Locus	213	1.5e-13	
Protein name  Description	712	5934	Length	Length 2031 Locus	213 S Name	1.5e-13 <u>Acc#</u>	
22289781 <u>f3</u> 195 Protein name	712	5934	Length 676	Length  2031  Locus  sp:PLE	213 s Name C_CAUCR	1.5e-13 <u>Acc#</u> P37894	
Protein name  Description	712	5934	Length	Length 2031 Locus	213 S Name	1.5e-13 <u>Acc#</u>	
Protein name  Description  NON-MOTILE AND PHACE	712  GE-RESISTA  NTID	5934	Length 676 EIN, NT Length	Length  Locus  Sp:PLE	213 s Name C_CAUCR	1.5e-13 <u>Acc#</u> P37894	
Protein name  Description  NON-MOTILE AND PHACE  ORF Name	712  GE-RESISTA  NTID	5934 ANCE PROTE	Length  676  EIN,  NT  Length	Length  Locus  Sp:PLE  AA  Length	213 s Name C_CAUCR	1.5e-13 <u>Acc#</u> P37894	
Protein name  Description  NON-MOTILE AND PHACE  ORF Name  22902267_c2_318	712  GE-RESISTA  NTID	5934  ANCE PROTE	Length  676  EIN,  NT  Length	Length  Locus  Sp:PLE  AA  Length	213 S Name C_CAUCR Score	1.5e-13 Acc# P37894 Probability	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
23477187_f1_54	714	<u>5936</u>		948	398	] 5.9e-37	
Protein name			L L	Locu	s Name	Acc#	
BrkB		<u> </u>		pir:I4		I40328	
Description				I (		<del></del>	
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
23522812 <u></u> £2 <u></u> 70	715	5937	63 1	.92			
Protein name				Locus	s Name	Acc#	
Description							
NO-HIT							
			3700	7.7		×	
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
23593750 <u></u> £2 <u></u> 84	716	5938	726 2	181			
Protein name				Locus	Name	Acc#	
Description						e 10	
NO-HIT			*				7
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
2381338 <u></u> c2 <u></u> 316	717	5939	347 1	.044	87	0.025	
Protein name		7		Locus	Name	Acc#	
hypothetical prote	ein PH0161	*		pir:G7	L237	G71237	
Description							
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability	
24259427 <u></u> c3381	718	5940	265 7	98			
Protein name		16		Locus	Name	Acc#	
Description							
NO-HIT							1
							J

ODE Namo	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name [24335943 cl 236	7 719	5941		158	190	8.0e-12	<b>—</b>
		3341			Name	Acc#	<u></u> j
Protein name	-ianl nwotoi	ъ МПЦООЛ				B69218	
conserved hypothet		.11 M11004		pir:B69	,718	B69218	
Description							
¥ 2			NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length	-		
24408517 <u>c3</u> 384	720	5942	93 2	82			
<u>Protein name</u>				Locus	Name	Acc#	
Description							
NO-HIT					· 8 1		]
			NT	AA	0	Dachahilitar	*
ORF Name	NTID	AAID	Length	Length	Score	Probability	
24640677 <u></u> c2 <u></u> 317	721	5943	290	373			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							1
			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
24641080 <u></u> c1 <u></u> 214	722	5944	136				
Protein name			2.5	Locus	s Name	Acc#	
Description							
NO-HIT							7
			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
24643887 <u></u> c2 <u></u> 314	723	5945	355	1068			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT					_		]

ř.	HV.
***	120
į,	1171
Ÿ	H
4	
	pat
97	DESI SERI
ST. ST.	186 186 186 186 186 186 186 186 186 186
Henry	
H	Hall Hall
4	177
	NUL NUL
Mary H.	
100 H. 100 H	
100 H. 100 H	
100 H. 100 H	日本本 長日 15.4万
HE THAT IS NOT IS	日 日 元 元 日 日 元 日 日 日 日 日 日 日 日 日 日 日 日 日
HE THAT IS THAT IS THAT	和
HE THAT IS NOT IS	日 日 元 元 日 日 元 日 日 日 日 日 日 日 日 日 日 日 日 日

RF Name NTID AAID Length Length
4648538_c1_233
Protein name Locus Name Acc#
sp:Y746_METJA Q58156
Description
HYPOTHETICAL PROTEIN MJ0746
NT <u>AA</u> ORF Name NTID AAID Length Length
4650302_f1_16
Protein name  Description
O-HIT
NT AA Score Probability
ORF Name NTID AAID Length Length
NT AA Score Probability  NTID AAID Length Length  24650912c3349
NRF Name         NTID         AAID         Length         Length         Score         Probability           24650912_c3_349         726         5948         519         1560         1415         1.0e-144           Protein name         Locus Name         Acc#
NT AA Score Probability  24650912_c3_349
NRF Name         NTID         AAID         Length         Length         Score         Probability           24650912_c3_349         726         5948         519         1560         1415         1.0e-144           Protein name         Locus Name         Acc#
NT Name         NTID         AAID         Length         Length         Score         Probability           24650912_c3_349
NTID AAID Length Length Score Probability  AAA Score Probability  AAA550912_c3_349
NTID AAID Length Length Score Probability  AAA Score Probability  AAA550912_c3_349
DRF Name         NTID         AAID         Length         Length         Score         Probability           24650.912_c3_349         726         5948         519         1560         1415         1.0e-144           Protein name         Locus Name         Acc#           sodium/proline symporter (proline permease)         pir:C69115         C69115           Description         NT AA         Score         Probability           ORF Name         NTID AAID Length Length         Length         Score         Probability           247.98457_t3_162         727         5949         500         1503

		NT AA Score Probability	
ORF Name	NTID AAID	Length Length	
24806567_c1_221	728 5950	670 2013 233 1.2e-21	
Protein name		Locus Name Acc#	
		sp:DSBD_HAEIN P44919	
Description			
BIOGENESIS PROTEIN	CYCZ)		
		277	
ORF Name	NTID AAID	NT AA Score Probability Length Length	
24853385 c2 321	729   5951	343 1032	
Protein name	,	Locus Name Acc#	
Description			1
NO-HIT			
		NT AA Garas Dachahilita	
ORF Name	NTID AAID	Length Length Score Probability	
2504787 <u>f3</u> 189	730 5952	258 777 89 0.0093	
Protein name		Locus Name Acc#	
ORF128 hypothetica	I protein	gp:AF008210 AF008210	
Description			
Buchnera aphidicol	a genomic fragment	containing (chaperone Hsp60)groEL, DNA	
biosynthesis initia	ting protein (dnaA)	), ATP operon(atpCDGAHFEB), and	
putative chromosome	replication protei	in (gidA)genes, complete cds; and	
termination factor	Rho (rho) gene, par	rtialcds.	İ
2.0		NT AA Garage Duchehilita	
ORF Name	NTID AAID	Length Length Score Probability	
25859425 <u></u> c3 <u></u> 373	731 5953	64 195 92 0.00016	
Protein name	1	Locus Name Acc#	
hypothetical prote	in ssr1765	pir:S74779 S74779	
Description	8		

ORF Name NTID AAID Length Length Score Probability	
26220277_c1_252	
Protein name Locus Name Acc#	
Description NO-HIT	
NO-HII	
ORF Name NTID AAID Length Length Probability	
2636.7.943£2120	
Protein name Locus Name Acc#	
Description	
NO-HIT	
ORF Name NTID AAID Length Length	
26369087 <u></u> c2 <u>299</u>	]
Protein name Locus Name Acc#	
sp:YHAI_ECOLI P42622	
Description	
Description HYPOTHETICAL 13.5 KD PROTEIN IN EXUR-TDCC INTERGENIC REGION	
Description  HYPOTHETICAL 13.5 KD PROTEIN IN EXUR-TDCC INTERGENIC REGION  NT AA Score Probability  ORF Name NTID AAID Length Length	
Description  HYPOTHETICAL 13.5 KD PROTEIN IN EXUR-TDCC INTERGENIC REGION  ORF Name  NTID  AAID Length Length Score Probability  26461627cl223	
Description           HYPOTHETICAL 13.5 KD PROTEIN IN EXUR-TDCC INTERGENIC REGION           ORF Name         NTID         AAID         Length         Score         Probability           26.46.16.27cl223	_
Description  HYPOTHETICAL 13.5 KD PROTEIN IN EXUR-TDCC INTERGENIC REGION  ORF Name  NTID AAID Length Length Length 500 4.0e-57	
Description           HYPOTHETICAL 13.5 KD PROTEIN IN EXUR-TDCC INTERGENIC REGION           ORF Name         NTID         AAID         Length         Score         Probability           26.46.16.27cl223	<del></del>

ORF Name	NTID AAID	NT AA Score Probabili Length Length	ty
2928387 c1_215	736   5958	197   594	
Protein name		Locus Name Acc#	
Description			
NO-HIT			
		NT AA Garra Probabili	
ORF Name	NTID AAID	NT AA Score Probabili Length Length	СY
29314080 <u></u> c1231	737 5959	724 2175 1417 6.1e-14	5
Protein name		Locus Name Acc	<u> </u>
		sp:DCP_ECOLI	
Description		<u>«</u>	
PEPTIDYL-DIPEPTIDAS	SE DCP, (DIPEPTIDYI	CARBOXYPEPTIDASE)	
		NT AA Score Probabili	.ty
ORF Name	NTID AAID		
3.0256.552f140	738 5960	243 730 272	
Protein name		Locus Name Acc	
uridine kinase udk		pir:G69728 G697	28
Description			
1		NT AA Score Probabil:	ity
ORF Name	NTID AAID	Length Length	
32228408 <u>f3</u> 176	. 739 5961	464 1395 473 6.6e-4!	<b>)</b>
Protein name		Locus Name Acc	
unknown		gp:AF086638 AF08	6638
Description	2		
Pseudomonas putida	CumA precursor (c	umA) and CumB (cumB) genes,complete	E -
cds; and unknown ge	ncs.		

ORF Name	NTID A	NT AID Length	AA Length Scor	e <u>Probability</u>	
33235905_c3_385	740 596	2 476	1431		
Protein name			Locus Name	Acc#	
Description					_
NO-HIT					]
ORF Name	NTID A	<u>NT</u> AID <u>Leng</u> th	AA Length Scor	e Probability	
34183438 <u></u> £1 <u></u> 35	741 596	3 239	720 430	2.4e-40	
Protein name			Locus Name	Acc#	
			sp:YHHW_ECOL	P46852	
Description		- AC JACO - A A A	15 APRIL A 17 A 7 A 1	7.50.0	1
HYPOTHETICAL 26.3	KD PROTEIN IN	GNTR-GGT INTE	RGENIC REGION	(F231)	
ORF Name	NTID A	AID Length	AA Length Scor	e Probability	
34625053 <u></u> f2 <u></u> 129	742 596	4 493	1482 352	1.6e-34	- ·
Protein name			Locus Name	Acc#	
damage-inducible	protein PAB0243		pir:A75151	A75151	
Description					
ORF Name	NTID A	<u>NT</u> AID Length	AA Length	e Probability	5
36056510 <u></u> .c1 <u></u> 251	743 596	5 422	1269		
Protein name					
			Locus Name	Acc#	
Description	$^* \cdot \cdot_{x_k}$		Locus Name	Acc#	
Description			Locus Name	Acc#	]
		<u>NT</u>	AA Scor		]
NO-HIT ORF Name		AID Length	AA Length Scor	e Probability	
NO-HIT  ORF Name  36132912_c2_281		AID Length	AA Scor	e Probability	
NO-HIT  ORF Name  36132912_c2_281  Protein name	744 596	AID Length	AA Scor Length 106 Locus Name	e Probability	
NO-HIT  ORF Name  36132912_c2_281	744 596	AID Length	AA Length Scor	e Probability	





ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
3912781_f2_98	745	5967	338	.017	317	1.8e-37
Protein name				Locus	s Name	Acc#
hypothetical protein protein F14M4.29:hyp	n F19D11.1 oothetical	6:hypothe protein 1	etical F14M4.29	pir:To:	2689	
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
3.928.550 <u></u> £3 <u></u> 16.1	. 746	5968	1054	3165	325	4.3e-45
Protein name				Locu	s Name	Acc#
115K outer membrane protein	e protein p	orecursor	:SusC	pir:JC	6027	JC6027
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 3.938.831f3147	<u>NTID</u>	<u>AAID</u>	Length		Score	Probability
		*	<u>Length</u>	Length		
3.938831 <u>f3147</u>		*	<u>Length</u>	Length 1251 Locu	1961	1.4e-202
3.938831_f3_147  Protein name  Description	. 747	5969	Length 416	Length 1251 Locu	1961 s Name	1.4e-202 <u>Acc#</u>
3.938831f3147 Protein name	. 747	5969	Length 416	Length 1251 Locu	1961 s Name	1.4e-202 <u>Acc#</u>
3.938831_f3_147  Protein name  Description	. 747	5969	Length 416	Length 1251 Locu	1961 s Name	1.4e-202 <u>Acc#</u>
Protein name  Description  CHONDRO-6-SULFATAS  ORF Name	747  E REGULATO  NTID	5969 RY PROTEI	Length  416  N  NT  Length	Length Locu sp:CHU	1961 s Name R_BACTN	1.4e-202 Acc# Q02550
3.938831f3147  Protein name  Description  CHONDRO-6-SULFATAS	747  E REGULATO  NTID	5969  RY PROTEI	Length  416  N  NT  Length	Length  Locu  Sp:CHU  AA  Length	1961 s Name R_BACTN Score	Acc# Q02550  Probability
Description  CHONDRO-6-SULFATAS  ORF Name	747  E REGULATO  NTID	5969  RY PROTEI	Length  416  N  NT  Length	Length  Locu  Sp:CHU  AA  Length  615	1961 s Name R_BACTN Score	1.4e-202  Acc# Q02550  Probability  2.6e-34

		71 71 TT	NT	<u>AA</u> Length	Score	Probability
ORF Name	NTID	AAID	Length	49	141	] [2.7e-09
4022312_c1_209	749	5971	282 8			
Protein name				Locus	Name	Acc#
ferredoxin (fdx-3)	homolog			pir:C69	294	C69294
Description						
			NT	AA		Parala de 2 7 de sa
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability
4069137 <u></u> c3 <u></u> 387	750	5972	301	006	127	1.9e-14
Protein name				Locus	Name	Acc#
leader peptidase Le	ep			gp:AF18	38620	AF188620
Description						
Bordetella pertuss:	is lep ope	ron, comp	lete sequ	ience.	-	2
			NT	AA		Dechabilite
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 4117150_f3_197		<u>AAID</u> 5973	<u>Length</u>		Score	Probability  9.4e-117
			Length	Length		
<u>4117150f3197</u>			Length	Length L278 Locus	1151	9.4e-117
<u>4117150f3197</u>			Length	Length L278 Locus	1151 S Name	9.4e-117 <u>Acc#</u>
4117150_f3_197 Protein name	751	5973	Length 426	Length Locus Sp:SR54	1151 S Name A BACSU	9.4e-117 <u>Acc#</u>
Protein name  Description	751	5973	Length 426	Length Locus sp:SR54	1151 S Name A BACSU	9.4e-117 <u>Acc#</u> P37105
Protein name  Description	751	5973	Length 426	Length Locus Sp:SR54	1151 S Name A BACSU	9.4e-117 <u>Acc#</u>
Protein name  Description  SIGNAL RECOGNITION  ORF Name	751  PARTICLE  NTID	5973 PROTEIN (	Length  426  FIFTY-FOLK  NT  Length	Length Locus Sp:SR54 JR HOMOLO	1151 S Name 4_BACSU OG)	9.4e-117 <u>Acc#</u> P37105
Protein name  Description  SIGNAL RECOGNITION	751  PARTICLE  NTID	PROTEIN (	Length  426  FIFTY-FOUNT  NT  Length	Length Locus Sp:SR54 UR HOMOLO  AA Length	1151 S Name  4_BACSU  OG)  Score	9.4e-117  Acc#  P37105  Probability
Protein name  Description  SIGNAL RECOGNITION  ORF Name  4461562_c3370	PARTICLE  NTID  752	PROTEIN (  AAID  5974	Length  426  FIFTY-FOUNT  NT  Length	Length Locus Sp:SR54 UR HOMOLO  AA Length	1151 S Name  4 BACSU  OG) Score  216 S Name	9.4e-117  Acc#  P37105  Probability  1.1e-17

ORF Name	NTID AAID	NT AA Length Length	Probability
4689092_c1_232	753 5975	199 600 244	1.2e-20
Protein name		Locus Name	Acc#
ferric uptake r	egulator homolog	gp:AF095596	AF095596
Description			
Staphylococcus gene, complete c		ric uptake regulator homo	olog (furB)
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
4698432_f2_121	754 5976	1023 687	1.4e-67
Protein name		Locus Name	Acc#
synthase III		pir:F70394	F70394
Description			
		NT AA G	
ORF Name	NTID AAID	Length Length Score	Probability
4734638 <u></u> c3 <u></u> 3.54		287 864 639	1.7e-62
Protein name		Locus Name	Acc#
		sp:HIS1_SALTY	P00499
Description			
ATP PHOSPHORIBO	SYLTRANSFERASE,		
		NT AA C	
ORF Name	NTID AAID	Length Length Score	Probability
4876563 <u></u> c1 <u></u> 216		153 462 361	4.9e-33
Protein name		Locus Name	Acc#
		sp:SMPB_BACSU	032230
Description		- '	
SMALL PROTEIN E	B HOMOLOG		

NT AA Score Probability NTID AAID Length Length	
DRF Name 1122 1200 1 8e-26	٦
1960812_f3_153	٦
Protein name Locus Name Acc#	
sp:THIO_BORBU O51088	
Description	
THIOREDOXIN (TRX)	
NT AA Score Probability	نصيب
ORF Name NTID AAID Length Length Score Probability	
5175875_c2_320	
Protein name Locus Name Acc#	
Description	
Description	
NO-HIT	
NT AA Score Probability	
ORF Name NTID AAID Length Length	
[502] [485] [1458] [1360] [6.7e-139]	
5313952 <u></u> £3 <u></u> 190	
5313.952f.3190	
5.3.1.3.9.5.2±.319.0	
5313.952t3190	
Protein name  Locus Name  Acc#  raw starch digesting amylase precursor  gp:AF067653  AF067653	
Protein name  Protein name  Locus Name  Acc#  raw starch digesting amylase precursor  Description  Cytophaga sp. raw starch digesting amylase precursor, gene, complete cds.	
Protein name  Protein name  Locus Name  Acc#  Taw starch digesting amylase precursor  Description  Description	
Protein name  Protein name  Locus Name  Acc#  raw starch digesting amylase precursor  Description  Cytophaga sp. raw starch digesting amylase precursor, gene, complete cds.  ORF Name  NTID  AAID  Length  Length  Score  Probability  Protein name  Acc#  AF067653  AF067653	
Protein name  Protein name  Locus Name  Acc#  Aro67653  Description  Cytophaga sp. raw starch digesting amylase precursor, gene, complete cds.  ORF Name  NTID  AAID  AAID  Length Length Length Length  759  Acc#  Aro67653  AF067653  Description  ORF Name  NTID  AAID  AAID  AAID  AAA  Length Length  AAA  242  2.0e-20	
Protein name         Locus Name         Acc#           raw starch digesting amylase precursor         gp:AF067653         AF067653           Description         Cytophaga sp. raw starch digesting amylase precursor, gene, complete cds.           ORF Name         NTID         AAID         NTID         AAID         Length         Length         Score         Probability           5795.7713152         760         5982         158         477         242         2.0e-20           Protein name         Locus Name         Acc#	
Protein name         Locus Name         Acc#           raw starch digesting amylase precursor         gp:AF067653         AF067653           Description         Cytophaga sp. raw starch digesting amylase precursor, gene, complete cds.           ORF Name         NTID         AAID         Length         Score         Probability           579577_±3_152	
Protein name         Locus Name         Acc#           raw starch digesting amylase precursor         gp:AF067653         AF067653           Description         Cytophaga sp. raw starch digesting amylase precursor, gene, complete cds.           ORF Name         NTID         AAID         NTID         AAID         Length         Length         Score         Probability           5795.7713152         760         5982         158         477         242         2.0e-20           Protein name         Locus Name         Acc#	

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
6056552_c3_386  Protein name	761 59	983	337 1	014 Locus	Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
6102312_c1_250 Protein name	. 762 5	984	349 1	Locus	Name	Acc#	
Description							-
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 6250003c1247		<u>AAID</u> 985	Length		Score	Probability 4.9e-05	
6250003_c1_247 Protein name	763 5	985	Length	Length L797 Locus	129 Name	4.9e-05 Acc#	
Protein name  conserved hypothet	763 5	985	Length	Length	129 Name	4.9e-05	1
6250003_c1_247 Protein name	763 5	985	Length	Length L797 Locus	129 Name	4.9e-05 Acc#	<b>]</b>
Protein name  conserved hypothet	763 5	985	Length	Length L797 Locus	129 Name	4.9e-05 Acc#	,
Protein name  conserved hypothet  Description	ical protein	985 BB0195	Length  598  NT  Length	Length Locus Locus AA	129 S Name 0124	4.9e-05 Acc# C70124	
Protein name    Conserved hypothet	ical protein	985 BB0195	Length  598  NT Length	Length Locus  AA Length Locus Locus Locus	129 S Name 0124 Score 120 S Name	Acc# C70124  Probability  1.7e-07  Acc#	,
Protein name  Conserved hypothet  Description  ORF Name	ical protein	985 BB0195	Length  598  NT Length	Length Locus  AA Length Locus Locus Locus	129 S Name 0124 Score	4.9e-05 Acc# C70124  Probability 1.7e-07	

ORF Name 829692 c2 303	<u>NTID</u>	<u>AAID</u>	NT Length	AA Length	Score	Probability	
Protein name					s Name	Acc#	
Description							
NO-HIT							1
							_
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
9.06.9.03c33.46	766	5988	287	364			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							]
							-
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
978387_c1_248	767	5989	130	393			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT		-			0.	*	7
				7.7			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
9900327 <u>c3</u> 324	768	5990	237	714	149	1.5e-10	
Protein name				Locus	Name	Acc#	
hypothetical protein	n PH1670			pir:F71	1047	F71047	
Description						0	
			NT	AA		×	
ORF Name	NTID	AAID	<u>Length</u>	<u>AA</u> Length	Score	Probability	
11721040 <u>f1</u> 42	769	5991	78	237			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT						7	]
							<u>.                                    </u>

NTID   AAID   Length   AEC#   AF126471   AF1	NT AA Score Probability
Description	OK! Name   17.69-48
Man26A   gp:AF126471   AF126471	1256885_t3_133
Description	Protein name
NT	Man26A gp:AF126471 AF126471
ORF Name         NTID         AAID         Length         Length         Score         Probability           12773255_c2_289         771         5993         519         1560         417         5.7e-39           Protein name         Locus Name         Acc#           Conserved hypothetical protein         pir:B72391         B72391           Description         NTID         AAID         Length         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Acc#         Acc#           Protein name         Locus Name         Acc#         Acc#         Acc#         Acc#           AryIsulfatase precursor (EC 3.1.6.1)         Sp:D90791         Sp:D90791         Description         Acc#         Acc#           E.coli genomic DNA, Konara clone #280(33.7-34.1 min.)         NT         AA         Score         Probability           ORF Name         NTID         AAID         Length         Length         I.4e-09           Protein name         Locus Name         Acc#           TRK system potassium uptake protein (trkA)         Sp:U32745	
NTID   AAID   Length   Length   Score   Probability	Cellulomonas fimi Man26A (man26A) gene, complete cds.
NTID   AAID   Length   Locus Name   Acc#   Locus Name   Acc#   Locus Name   Acc#   Length   Lengt	NT AA Score Probability
Protein name         Locus Name         Acc#           Conserved hypothetical protein         pir:B72391         B72391           Description         NTID         AAID         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         6.8e-36           Protein name         Locus Name         Acc#           Arylsulfatase precursor (EC 3.1.6.1)         gp:D90791           Description         E.coli genomic DNA, Kohara clone #280(33.7-34.1 min.).           ORF Name         NTID         AAID         Length         Length         Score         Probability           13711437_cl_21_215         773         5995         106         321         148         1.4e-09           Protein name         Locus Name         Acc#           TRK system potassium uptake protein (trkA)         gp:U32745	Tongth Length
Description	12773255_c2_289
Description   NT	Protein name Locus Name Acc#
ORF Name         NTID         AAID         Length         Length         Score         Probability           13709635_c1_211         772         5994         522         1569         319         6.8e-36           Protein name         Locus Name         Acc#           AryIsuIfatase precursor (EC 3.1.6.1)         gp:D90791           Description         E.coli genomic DNA, Kohara clone #280(33.7-34.1 min.)           ORF Name         NTID         AAID         NTID         AAA Length         Score         Probability           13711437_c1_215         773         5995         106         321         148         1.4e-09           Protein name         Locus Name         Acc#           TRK system potassium uptake protein (trkA)         gp:U32745	conserved hypothetical protein pir:B72391 B72391
ORF Name         NTID         AAID         Length         Length         Score         Probability           13.70.96.35_c1_211	Description
ORF Name         NTID         AAID         Length         Length         Score         Probability           13.70.96.35_c1_211	NIII AA
Description	Score Probability
Arylsulfatase precursor (EC 3.1.6.1)   gp:D90791	13.70.9635_c1_211772
Description           E.coli genomic DNA, Kohara clone #280(33.7-34.1 min.).           ORF Name         NTID         AAID         NTID         AAID         Length         Length         Probability           13711437c1215	Protein name Locus Name Acc#
E.coli genomic DNA, Kohara clone #280(33.7-34.1 min.).    NT	Arylsulfatase precursor (EC 3.1.6.1) gp:D90791
ORF Name         NTID         AAID         Length         Length         Score         Probability           13.71143.7c1215	Description
ORF Name         NTID         AAID         Length         Length         Score         Floadsfift           13.7.1143.7c1215	E.coli genomic DNA, Kohara clone #280(33.7-34.1 min.).
ORF Name         NTID         AAID         Length         Length         Score         Floadsfift           13.7.1143.7c1215	NT AA Grana Probability
Protein name  Locus Name  Acc#  TRK system potassium uptake protein (trkA)  gp:U32745	Score Floodbilley
TRK system potassium uptake protein (trkA) gp:U32745	13711437c1215
	Protein name Locus Name Acc#
	TRK system potassium uptake protein (trkA) gp:U32745
Descripcion	Description
Haemophilus influenzae Rd section 60 of 163 of the complete genome.	Distriction (0 of 162 of the complete genome.

ORF Name         NTID         AAID         Length         Length         Score         Probability           14651512_f1_5         774         5996         469         1410         782         1.2e-77           Protein name         Locus Name         Acc#           sp:YAGG_ECOLI         sp:YAGG_ECOLI
Description
HYPOTHETICAL SYMPORTER IN PERR-ARGF INTERGENIC REGION
ORF Name         NTID         AAID         Length         Length         Score         Probability           14726062_c1_203         775         5997         664         1995           Protein name         Locus Name         Acc#
Description
NO-HIT
ORF Name         NTID         AAID         Length         Length         Score         Probability           166.0.032.7£313.0
dimethylamine corrinoid protein MtbC gp:AF102623 AF102623
Description
Methanosarcina barkeri dimethylamine corrinoid protein MtbC (mtbC), trimethylamine methyltransferase MttB (mttB), trimethylaminecorrinoid protein MttC (mttC), putative transmembrane protein MttP(mttP), and dimethylamine methyltransferase MtbB1 (mtbB1) genes, complete cds.
ORF Name NTID AAID Length Length Score Probability
ORF Name         NTID         AAID         Length         Length           20525252.c2307
Protein name Locus Name Acc#
p:NU5C_SYNP2 P31971 Description
NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 5,

ODE Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name		*		320	354	] [1.5e-31 ]
2117177_£2_71	778	6000	439 1	1320	334	
Protein name				Locus	Name	Acc#
endo-1,4-beta-manno	osidase			pir:D72	2278	D72278
Description						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
21664650 <u></u> ±1 <u></u> 6	. 779	6001	396 1	1191	239	6.0e-18
Protein name				Locus	Name	Acc#
renin-binding prote slr1975:protein slr		protein	:protein	pir:S7	649	S75649
Description				-		***
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 21875377f17	,	<u>AAID</u> 6002	Length	<del></del>	Score	Probability
,	,		Length	Length	Score S Name	Probability  Acc#
218.753.7.7£17	,		Length	Length		
21875377fl7	,		Length	Length		
Protein name  Description	,		Length	Length	s Name	Acc#
Protein name  Description	,		Length	Length  1809  Locus		
Protein name  Description  NO-HIT	780 NTID	6002	Length  602  NT Length	Length  Locus	s Name	Acc#
Protein name  Description  NO-HIT  ORF Name	780 NTID	AAID	Length  602  NT Length	Length Locus  AA Length	Score	Acc# Probability
Protein name  Description  NO-HIT  ORF Name	780 NTID	AAID	Length  602  NT Length	Length Locus  AA Length	Score  475 S Name	Acc# Probability 4.5e-44
Protein name  Description  NO-HIT  ORF Name  22270002_t2_70.	780 NTID	AAID	Length  602  NT Length	Length Locus  AA Length 777 Locus	Score  475 S Name	Acc#  Probability  4.5e-44  Acc#

ODE Nama	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score	Probability
ORF Name	7	398 11197	1 199	0.036
22550917_c2_317	782 0004		J L	Acc#
<u>Protein name</u>			us Name	Acc#
endo-beta-1,3-gluca	anase precursor	gp:AF	013169	
Description				- 1
Pyrococcus furiosu operon, complete se- cds; and 2-phosphog	quence; biotin liga	se BirA homolog	(birA) gen	;adh-lam e, complete
ORF Name	NTID AAID	NT AA Length Length	Score	Probability
23712837_f2_81	783 6005	377 1134	169	4.8e-12
Protein name		Loc	us Name	Acc#
conserved hypothet	ical protein SC9C7.	14c pir:T	35965	T35965
Description			-	
*	- 10	7.7		
ORF Name	NTID AAID	<u>NT AA</u> Length Lengtl	Score	Probability
2426.03.02 <u>f3</u> 13.4	784 6006	398 1197	283	9.0e-25
Protein name		Loc	us Name	Acc#
conserved hypothet	ical protein	pir:E	72278	B72278
Description				* * *
ORF Name	NTID AAID	NT AA Length Lengt	h Score	Probability
24409668 <u></u> £3 <u></u> 124	785 6007	140 423		
Protein name		Loc	us Name	Acc#
Description		*		

1	#3 <sup>1</sup>
	No. of Street, or other Persons
STREET S	25
91	Tight.
100	:22
Service of the last	
History	137
W. A.	821 11 1127
į,	Ti-
Ž,	175
W.	1176
	Hart House
*	The state of the s
Mary Mary	en en en en en en en en en en en en en e
Mary Mary	是 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
THE REAL PROPERTY.	The state of the s
THE PERSON OF THE PERSON OF	新 製 製 製 製 製 製 製 製 製 製 製 製 製 製 製 製 製 製 製
雅 机焊 医 珠山色 蓝 珠山色	Early 新 等 等 等 等 等 等 等 等 等
雅 机焊 医 珠山色 蓝 珠山色	新 製 製 製 製 製 製 製 製 製 製 製 製 製 製 製 製 製 製 製

			NT	<u>AA</u>	Score	Probability	
ORF Name	NTID	AAID	Length	Length			_
24415962_c2_304	786	6008	160	183	231	2.9e-19	
Protein name				Locus	Name	Acc#	
NADH dehydrogenase RP795	(ubiquinor	ne), I ch	ain I	pir:E7	.640	E71640	
Description							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24492137f1_2	787	6009	1075	3228	163	1.4e-07	
Protein name				Locus	s Name	Acc#	
probable secreted	glucosidas	ē		pir:T3	5164	T35164	
Description			0			ź	
* ×							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
		<u>AAID</u>	Length		Score	Probability	
246.427.87 <u></u> c2 <u></u> 316			Length	Length	Score s Name	Probability  Acc#	
24642787c2316 Protein name			Length	Length			
Protein name  Description			Length	Length			1
24642787c2316 Protein name			Length	Length			,
Protein name  Description			Length	Length			,
Protein name  Description  NO-HIT	788 NTID	6010	Length  426	Length  Locu  AA	s Name	Acc#	]
Protein name  Description  NO-HIT  ORF Name  246.440.52_c1_20.5	788 NTID	6010 AAID	Length  426  NT Length	Length  Locu  AA  Length	s Name	Acc# Probability	
Protein name  Description  NO-HIT  ORF Name	788 NTID 789	AAID	Length  426  NT Length	Length  Locu  AA  Length	Score 197 s Name	Acc# Probability 2.4e-26	
Protein name  Description  NO-HIT  ORF Name  246.440.52_c1_20.5	788 NTID 789	AAID	Length  426  NT Length	Length  Locu  AA  Length  1218  Locu	Score 197 s Name	Acc#  Probability  2.4e-26  Acc#	

ODE Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name				706	176	3.8e-08
24645437_c3_384	790 60	012	901 2	706	170	
Protein name				Locus	Name	Acc#
115K outer membrane	protein pre	cursor:	SusC	pir:JC6	027	JC6027
protein						
Description						:
			NT	AA		
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability
24648388t3198	791 60	013	239 7	17	518	1.1e-49
Protein name				Locus	Name	Acc#
probable glycosyl h	ıydrolase		e.	pir:T36	467	T36467
Description						
	- 8					
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
25400260 <u>c1</u> 220	. 792 60	014	279 8	40	568	5.7e-55
Protein name				Locus	s Name	Acc#
				sp:NUO	H_ECOLI	
Description				- 10	F .	100
OXIDOREDUCTASE CHA	IN 8) (NUO8)	×				
			NT	AA	0	Drobability
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 25817656 <u></u> c3. <u></u> 368		<u>AAID</u>	Length	<del></del>	<u>Score</u> 871	Probability 4.4e-87
			Length	Length	-	
25817656 <u></u> 23 <u></u> 368			Length	Length  511  Locu	871	4.4e-87
25817656 <u></u> 23 <u></u> 368			Length	Length  511  Locu	871 s Name	4.4e-87

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26230265_f2_67  Protein name	794	6016	625	Locus	s Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
26360717 <u>f3</u> 183	. 795	6017	219	660	346	1.9e-31
Protein name				Locus	s Name	Acc#
phosphoglycolate pl	nosphatase	gph) ho	molog	pir:C7	0184	C70184
Description						
					<u> </u>	
			NT	$\underline{AA}$	Gaore	Probability
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 26366542_c2_308		<u>AAID</u>			Score	Probability 1.9e-73
			Length	Length	*	
263.6.6.542c23.0.8	796 (ubiquinc	6018	Length 498	Length	719 s Name	1.9e-73
26366542_c2_308 Protein name  NADH dehydrogenase	796 (ubiquinc	6018	Length 498	Length  1497  Locu	719 s Name	1.9e-73 <u>Acc#</u>
Protein name  NADH dehydrogenase 4.2:protein slr1291	796 (ubiquinc	6018	Length 498	Length  Locu  pir:S7	719 s Name	1.9e-73 <u>Acc#</u> S74687  Probability
Protein name  NADH dehydrogenase 4.2:protein slr1291  Description	(ubiquino :protein s	one), chai	Length 498 n	Length  Locu  pir:S7	719 s Name 4687	1.9e-73 <u>Acc#</u> S74687
Protein name  NADH dehydrogenase 4.2:protein slr1291  Description  ORF Name	(ubiquino :protein s	one), chai	Length  498  n  NT  Length	Length  Locu  pir:S7  AA  Length  1596	719 s Name 4687 Score	1.9e-73 <u>Acc#</u> S74687  Probability
Protein name  NADH dehydrogenase 4.2:protein slr1291  Description  ORF Name	796  (ubiquino :protein s	AAID	Length 498  n  NT Length	Length  Locu  pir:S7  AA  Length  1596	719 s Name 4687 Score 738 s Name	1.9e-73  Acc#  S74687  Probability  4.7e-85

		NT AA Score Probability
ORF Name	NTID AAID	<u>Length</u> <u>Length</u> 3.8e-13
26587708_t2_65	798 6020	
Protein name		Locus Name Acc#
unknown		gp:U96771 U96771
Description		
Prevotella bryan mannanase genes,	tii putative polygala complete cds; and unl	ncturonase,B-1,4-endoglucanase, and mowngenes.
		NT AA Score Probability
ORF Name	NTID AAID	Length Length
26594137_f2_75	799 6021	336 1011 293 7.9e-26
Protein name		Locus Name Acc#
methylcobamide:Co	oM methyltransferase	isozyme gp:AF013713 AF013713
Description		
(mtbA) monomethy	lamine corrinoid pro thyltransferase (mtm	B), putative monomethylamine permease
(memi ) , and amino.	n genes, comprete cu	
ORF Name	NTID AAID	NT AA Score Probability Length Length
	NTID AAID	NT AA Score Probability
ORF Name	NTID AAID	NT AA Score Probability
ORF Name 26604712_c2_309	NTID AAID	NT AA Score Probability Length Length 725 1.3e-71
ORF Name 26604712_c2_309	NTID AAID	NT AA Length Length Score Probability  485   1458   725   1.3e-71  Locus Name Acc#
ORF Name  26.6.04712_c2_309  Protein name  Description	NTID AAID	NT Length         AA Length         Score         Probability           485         1458         725         1.3e-71           Locus Name         Acc#           sp:NU2C_SYNY3         P72714
ORF Name  26.6.04712_c2_309  Protein name  Description	NTID AAID	NT AA Score Probability  Length Length 725 1.3e-71  Locus Name Acc#  sp:NU2C_SYNY3 P72714  AIN 2,
ORF Name  26.6.04712_c2_309  Protein name  Description	NTID AAID	NT Length         AA Length         Score         Probability           485         1458         725         1.3e-71           Locus Name         Acc#           sp:NU2C_SYNY3         P72714
ORF Name  26.6.0.4.7.12	NTID AAID  800 6022  NE OXIDOREDUCTASE CH	NT AA Score Probability  Length Length  1485  Locus Name Acc#  Sp:NU2C_SYNY3  P72714  AIN 2,  NT AA Score Probability
ORF Name  26.6.047.12c23.09  Protein name  Description  NADH-PLASTOQUING  ORF Name	NTID AAID  800 6022  NE OXIDOREDUCTASE CH	NT AA Score Probability  Length Length  725  Locus Name Acc#  Sp:NU2C_SYNY3  P72714  AIN 2,  NT AA Score Probability  Length Length Score Probability
ORF Name  26.6.04.7.12c23.0.9  Protein name  Description  NADH-PLASTOQUING  ORF Name  29.45.75.5.7c23.0.0	NTID AAID  800 6022  NE OXIDOREDUCTASE CH	NT Length         AA Length         Score         Probability           485         1458         725         1.3e-71           Locus Name         Acc#           sp:NU2C_SYNY3         P72714           AIN 2,         NT AA Score         Probability           Length         Length         225         1.3e-18
ORF Name  26.6.04.7.12c23.0.9  Protein name  Description  NADH-PLASTOQUING  ORF Name  29.45.75.5.7c23.0.0	NTID AAID  800 6022  NE OXIDOREDUCTASE CH	NT Length         AA Length         Score         Probability           485         1458         725         1.3e-71           Locus Name         Acc#           sp:NU2C_SYNY3         P72714           AIN 2,         NT Length Length         Score         Probability           Length Length         225         1.3e-18           Locus Name         Acc#

ODE Name	NTID AAID	<u>NT</u> <u>AA</u> Length Lengt	Score	Probability
ORF Name	144		- 7 [163	4.7e-12
31776708_c3_371	802 6024	65 198		
Protein name			us Name	Acc#
NADH dehydrogenase	(ubiquinone), I	chain nuoB pir:(	70413	C70413
Description				
ORF Name	NTID AAII	<u>NT</u> <u>AA</u> Length Lengt	h Score	Probability
32532838 <u>f2</u> 80	803 6025	518 1557		
Protein name		Loc	cus Name	Acc#
Description				
NO-HIT				
		NT AA		
ORF Name	NTID AAII	<del></del> .	h Score	Probability
33835790 <u></u> t2 <u></u> 74	804 6026	251 756	92	0.045
Protein name		Lo	cus Name	Acc#
hypothetical prote	ein	pir:	272397	C72397
Description		1		
• •		7.7		
ORF Name	NTID AAII	NT AA Length Lengt	h Score	Probability
3.6132686 <u></u> c336.4	805 6027	114 345		
Protein name		<u>Lo</u>	cus Name	Acc#
Description				
NO-HIT				
		NT AA	Score	Probability
ORF Name	NTID AAI	<u> Length Lengt</u>	<u></u>	
36360962 <u></u> c2305	806 6028	172 519	204	2.1e-16
Protein name		T <sub>1</sub> O	cus Name	Acc#
NADH dehydrogenas	e (ubiquinone), I		C71839	C71839

		NT	AA Score	Probability
ORF Name	NTID AAID		ength ——	
3907132_f2_90	807 6029	932 27	99 470	7.6e-41
Protein name			Locus Name	Acc#
sensory transduct	ion histidine kinas		pir:S75130	S75130
slr2098:protein sl	lr2098:protein slr20	198		
Description				
		NT	AA Score	Probability
ORF Name	NTID AAID	<u>Length</u>	Length Source	1
39953.02 <u></u> c2 <u></u> 3.01	808 6030	185 55	8 319	1.4e-28
Protein name			Locus Name	Acc#
NADH dehydrogena	se I, subunit nuoB		gp:ECNUOO	X68301
Description				
E.coli DNA seque	nce of nuo operon.			
			_	
		NT	ΔΔ	
ORF Name	NTID AAID	<u>NT</u> Length	AA Length Score	Probability
ORF Name 4146907c1202		Length	Score	Probability  2.4e-81
		Length	Length Score	
4146907 <u></u> c1 <u></u> 202	809 6031	Length	Length Score 714	2.4e-81
4146907 <u></u> c1 <u></u> 202	809 6031	Length	Length Score 714  Locus Name	2.4e-81 <u>Acc#</u>
Protein name receptor antigen Description	(RagA)	<u>Length</u>	Length Score Tocus Name  Gp:PGI130872	2.4e-81 Acc# AJ130872
Protein name receptor antigen Description	(RagA)	<u>Length</u>	Length Score Tocus Name  Gp:PGI130872	2.4e-81 Acc# AJ130872
Protein name receptor antigen Description Porphyromonas gi	(RagA)	Length  [1071] [32]  or antigen	Length Score  Tocus Name  Gp:PGI130872  (rag) locus enc	2.4e-81 Acc# AJ130872 codinga major
Protein name receptor antigen Description Porphyromonas gi	(RagA)	Length  [1071] 32  or antigen	Length Score  Length 714  Locus Name  Gp:PGI130872  (rag) locus enc	2.4e-81 Acc# AJ130872
Protein name receptor antigen Description Porphyromonas gi immunodominant 55	RagA)  ngivalis W50 recept SkDa antigen.  NTID AAID	Length  1071 32  or antigen  NT Length	Length  Tocus Name  Ep:PGI130872  (rag) locus enc	2.4e-81 Acc# AJ130872 codinga major
Protein name receptor antigen Description Porphyromonas gi immunodominant 55	RagA)  ngivalis W50 recept SkDa antigen.  NTID AAID	Length  1071 32  or antigen  NT Length	Length  Tocus Name  Ep:PGI130872  (rag) locus end  AA Score	2.4e-81  Acc#  AJ130872  codinga major  Probability
Protein name receptor antigen Description Porphyromonas gi immunodominant 55  ORF Name	RagA)  ngivalis W50 recept SkDa antigen.  NTID AAID	Length  1071 32  or antigen  NT Length	Length Score  Length 714  Locus Name  Gp:PGI130872  (rag) locus end  AA Score  Length Score  173	2.4e-81  Acc#  AJ130872  codinga major  Probability  6.3e-18

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4566876 C2 285		6033		464	411	2.5e-38
				Locus	Name	Acc#
Protein name						P31447
				sp: YID	_ECOLI	
Description					5 H A 7 A 1	
HYPOTHETICAL 57.3	KD PROTEIN	IN EMRD-C	ELVG INTI	ERGENIC I	REGION	
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	1	
4975313_c3_369	812	6034	128	387		
Protein name				Locu	s Name	Acc#
Description						
		×				
NO-HIT			*			
	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name	MIID	11111				
		ICOSE ]	105 1	318	231	2.9e-19
5111038 <u></u> c2 <u></u> 306	813	6035	105	318	231	2.9e-19
5111038c2306 Protein name	<u>813</u>	6035	105	Locu	s Name	Acc#
	813	6035	105	Locu		
	813	6035	105	Locu	s Name	Acc#
Protein name				Locu	s Name	Acc#
Protein name  Description				Locu	s Name	Acc# Q00244
Protein name  Description			IN 4L,	Locu sp:NUL	s Name	Acc#
Protein name  Description  NADH-PLASTOQUINON	NE OXIDOREDU NTID	CTASE CHA	IN 4L,  NT  Length	Locu sp:NUL	s Name	Acc# Q00244
Protein name  Description  NADH-PLASTOQUINON  ORF Name	NE OXIDOREDU NTID	CTASE CHA	IN 4L,  NT  Length	Locu sp:NUL AA Length	s Name C_PLEBO Score	Acc# Q00244 Probability
Protein name  Description  NADH-PLASTOQUINON  ORF Name  5860027t3196	NE OXIDOREDUC  NTID	CTASE CHA  AAID  6036	IN 4L,  NT  Length	Locu sp:NUL AA Length	s Name C_PLEBO Score 519 as Name	Acc# Q00244  Probability 4.9e-46
Protein name  Description  NADH-PLASTOQUINON  ORF Name  586.0.027t3196	NE OXIDOREDUC  NTID	CTASE CHA  AAID  6036	IN 4L,  NT  Length	Locu sp:NUL AA Length 4143	s Name C_PLEBO Score 519 as Name	Acc# Q00244  Probability  4.9e-46  Acc#
Protein name  Description  NADH-PLASTOQUINON  ORF Name  586.0027t3196  Protein name  utilizing regulat  Description	NE OXIDOREDUC  NTID  814  Tory protein  a utilizing	CTASE CHA  AAID  6036  tutC  regulator	IN 4L,  NT Length  1380	Locu Sp:NUL  AA Length  4143  Locu  gp:TTU	S Name  C_PLEBO  Score  519  IS Name  157900  tutC),ut	Acc# Q00244  Probability  4.9e-46  Acc# U57900  ilizing
Protein name  Description  NADH-PLASTOQUINON  ORF Name  586.0027t3196	NE OXIDOREDUCE  NTID  814  tory protein  a utilizing  n tutB (tutB	AAID 6036 tutC regulator	IN 4L,  NT Length  [1380]  y protei ve DNA b	Locu sp:NUL  AA Length 4143  Locu gp:TTC n tutC (sindingpr	SCOTE  519  STORE  519  STORE  57900  tutC), ut	Acc# Q00244  Probability  4.9e-46  Acc# U57900  ilizing

		NTEL 7. 7.
ORF Name	NTID AAID	$rac{ ext{NT}}{ ext{Length}} rac{ ext{AA}}{ ext{Length}} rac{ ext{Score}}{ ext{Probability}}$
6444137_±3_129	815 6037	421 1266 138 2.8e-06
Protein name		Locus Name Acc#
CmuC protein		gp:MSP011317 AJ011317
Description		
Methylobacterium genes and genes e	sp. CM4, cobD, metF encoding Orf219 and C	F, cmuB, cmuC, partial cobC andcobQ, Drf361.
		NT AA Score Probability
ORF Name	NTID AAID	Length Length 1.4e-14
7074155_f1_1	816 6038	304 1133
Protein name		Locus Name Acc#
unknown		gp:096771 U96771
Description		
Prevotella bryar	itii putative polyga	lacturonase,B-1,4-endoglucanase, and
mannanase genes,	complete cds; and ur	nknowngenes.
mannanase genes,	complete cds; and ur	nknowngenes.
mannanase genes,	complete cds; and ur	NT AA Score Probability
mannanase genes,	omplete cds; and un	NT AA Score Probability
mannanase genes,	omplete cds; and un	NT AA Score Probability Length Length
mannanase genes,  ORF Name  7220153c2291	omplete cds; and un	NT AA Score Probability Length Length 1366 1.6e-139
mannanase genes,  ORF Name  7220153c2291	omplete cds; and un	NT AA Score Probability Length Length 1366 1.6e-139  Locus Name Acc#
ORF Name 7220153_c2_291 Protein name Description	omplete cds; and un	NT AA Score Probability Length Length 1366 1.6e-139  Locus Name Acc#  Sp:DXS_HAEIN P45205
ORF Name 7220153_c2_291 Protein name Description	NTID AAID  817 6039	NT AA Score Probability  Length Length 1366 1.6e-139  Locus Name Acc#  Sp:DXS_HAEIN P45205  E (DXP SYNTHASE)
ORF Name 7220153_c2_291 Protein name Description	NTID AAID  817 6039	NT AA Score Probability  Length Length 1366 1.6e-139  Locus Name Acc#  Sp:DXS_HAEIN P45205  E (DXP SYNTHASE)  NT AA Score Probability
ORF Name 7220153c2291 Protein name Description 1-DEOXYXYLULOSE.	NTID AAID  STID AAID  NTID AAID  AAID  AAID  NTID AAID	NT AA Score Probability  Length Length 1366 1.6e-139  Locus Name Acc#  Sp:DXS_HAEIN P45205  E (DXP SYNTHASE)  NT AA Score Probability
mannanase genes,  ORF Name  7220153_c2_291  Protein name  Description  1-DEOXYXYLULOSE  ORF Name  972167_c1_199	NTID AAID  STID AAID  NTID AAID  AAID  AAID  AAID  NTID AAID	NT AA Score Probability  Length Length 1366 1.6e-139  Locus Name Acc#  Sp:DXS_HAEIN P45205  E (DXP SYNTHASE)  NT AA Score Probability  Length Length Score Probability
ORF Name 7220153c2291 Protein name Description 1-DEOXYXYLULOSE.	NTID AAID  STID AAID  NTID AAID  AAID  AAID  AAID  NTID AAID	NT   AA   Score   Probability
mannanase genes,  ORF Name  7220153_c2_291  Protein name  Description  1-DEOXYXYLULOSE  ORF Name  972167_c1_199	NTID AAID  STID AAID  NTID AAID  AAID  AAID  AAID  NTID AAID	NT   AA   Score   Probability

		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
9954806_f1_3	819 6041	820 2463 1419	3.8e-145
Protein name		Locus Name	Acc#
beta-xylo-glucos:	idase	gp:TBZ56279	Z56279
Description			
T.brockii cglF,	cglG, xglS and cglT ge	enes.	
		NT AA	Dark hiliter
ORF Name	NTID AAID	Length Length Score	Probability
33212528 c3 11	820 6042	554 1662	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
**		NT AA Score	Probability
ORF Name	NTID AAID	NT <u>AA</u> Score Length Length	Probability
ORF Name 7119532f36	,	5000	Probability  0.017
	,	Length Length	
7119532 <u>f3</u> 6	,	Length Length 53	0.017
7119532 <u>f3</u> 6	,	Length Length 53  Locus Name	0.017 <u>Acc#</u>
7119532f36  Protein name  Description	<u>821</u> <u>6043</u>	Length Length 53  Locus Name  sp:GP38_CANFA	0.017 <u>Acc#</u> Q95152
7119532f36  Protein name  Description	<u>821</u> <u>6043</u>	Length Length Score  71 216 53  Locus Name  sp:GP38_CANFA  IN-TYPE MEMBRANE PROTEIN	0.017 <u>Acc#</u> Q95152
7119532f36  Protein name  Description	<u>821</u> <u>6043</u>	Length Length 53  Locus Name  sp:GP38_CANFA	0.017 <u>Acc#</u> Q95152
7119532_f3_6Protein name  Description  GLYCOPROTEIN 38	PRECURSOR (GP38) (MUC	Length Length 53  Locus Name  sp:GP38_CANFA  IN-TYPE MEMBRANE PROTEIN  NT AA Score	0.017 <u>Acc#</u> Q95152
7119532t36  Protein name  Description  GLYCOPROTEIN 38  ORF Name	PRECURSOR (GP38) (MUC	Length Length Score  71 216 53  Locus Name  sp:GP38_CANFA  IN-TYPE MEMBRANE PROTEIN  NT AA Length Length Score	0.017 <u>Acc#</u> Q95152
Protein name  Description  GLYCOPROTEIN 38  ORF Name  10007303t121	PRECURSOR (GP38) (MUC	Length Length Score  71 216 53  Locus Name  sp:GP38_CANFA  IN-TYPE MEMBRANE PROTEIN  NT AA Length Length  177 534	0.017  Acc#  Q95152  GP40)  Probability

** ***
F 43
77 65
10 mg
12 HS
(Zf 305)
10
1
# W.
æ
# # # # # # # # # # # # # # # # # # #
10 INC
in a
4400
EC DEC
717

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
10189501_±3_191	823	6045	81 ]	246		
Protein name		L	L  L	Locus	s Name	Acc#
						10011
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
10314088 <u>.</u> £1 <u>.</u> 46	824	6046	384	155	631	1.2e-61
Protein name	* 0			Locus	Name	Acc#
				sp:DINI		
Description					* ;	
Description	DIE DOMETN				*	
Description DNA-DAMAGE-INDUCI	BLE PROTEIN	P				
- 1.	BLE PROTEIN  NTID	P	NT Length	AA Length	Score	Probability
DNA-DAMAGE-INDUCI	NTID		Length		<u>Score</u>	Probability
DNA-DAMAGE-INDUCI	NTID	AAID	Length	Length		
DNA-DAMAGE-INDUCIORF Name	<u>NTID</u>	AAID	Length	Length	102 Name	1.4e-05
ORF Name  10548816f186  Protein name	<u>NTID</u>	AAID	Length	Length  89  Locus	102 Name	1.4e-05 <u>Acc#</u>
DNA-DAMAGE-INDUCION ORF Name  10548816tl86	<u>NTID</u>	AAID	Length	Length  89  Locus	102 Name	1.4e-05 <u>Acc#</u>
DNA-DAMAGE-INDUCION ORF Name  10548816tl86	<u>NTID</u>	AAID	Length	Length  89  Locus	102 Name	1.4e-05 <u>Acc#</u>
DNA-DAMAGE-INDUCT  ORF Name  10548816tl86  Protein name  hypothetical prot  Description	NTID 825 ein APE2457	<u>AAID</u> 6047	Length 62  NT Length	Length  89  Locus  pir:H72	102 Name 2476	1.4e-05 Acc# H72476
ORF Name  10548816tl86  Protein name hypothetical prot Description  ORF Name	NTID 825 ein APE2457	AAID 6047 AAID	Length 62  NT Length	Length  Locus  pir:H72  AA  Length  333	102 S Name 2476 Score	1.4e-05 Acc# H72476 Probability
DNA-DAMAGE-INDUCT  ORF Name  10548816t186  Protein name hypothetical prot  Description  ORF Name  1072177c2422	NTID 825 ein APE2457	AAID 6047 AAID	Length 62  NT Length	Length  Locus  pir:H72  AA  Length  333  Locus	102 S Name 2476 Score 671 Name	1.4e-05 Acc# H72476  Probability 3.2e-184
DNA-DAMAGE-INDUCT  ORF Name  10548816t186  Protein name hypothetical prot  Description  ORF Name  1072177c2422	NTID 825 ein APE2457	AAID 6047 AAID	Length 62  NT Length	Length  Locus  pir:H72  AA  Length  333	102 S Name 2476 Score 671 Name	1.4e-05 Acc# H72476  Probability 3.2e-184 Acc#

			NT	AA Tangth	Score	Probability	
ORF Name	NTID	AAID	Length	Length	10.05	1 9.2e-25	٦
11147938_f2_90	827	6049	401 1	.206	295		]
Protein name				Locus	Name	Acc#	
115K outer membra	ne protein p	recursor	:SusC	pir:JC	027	JC6027	
protein							
Description							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
1173557c1261	828	6050	349	L050	483	5.8e-46	]
Protein name				Locu	s Name	Acc#	
			4	sp:APB	E_HAEIN	P44550	
Description				=		ž.	
THIAMINE BIOSYNTE	HESIS LIPOPRO	OTEIN APB	E PRECURS	SOR			
I .							
			Mrti	ΔΔ			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	_
ORF Name		<u>AAID</u> 6051	Length		Score	Probability	_
			Length	Length	× ×		_
1180302 <u>f2</u> 93			Length	Length	256 s Name	1.3e-40	_
1180302 <u>f2</u> 93			Length	Length 1629 Locu	256 s Name	1.3e-40 <u>Acc#</u>	_
1180302 <u>f293</u>	 829		Length	Length  Locu  sp:STS	256 s Name	1.3e-40 <u>Acc#</u>	
Protein name  Description	 829	6051	Length  [542]  [SE C) (AS	Length  Locu  sp:STS	256 s Name	1.3e-40 Acc# P15589	
Protein name  Description	 829	6051	Length	Length  Locu  sp:STS	256 s Name	1.3e-40 <u>Acc#</u>	
Protein name  Description  SULFATE SULFOHYD  ORF Name	ROLASE) (ARY	6051	Length  542  E C) (ASC  NT  Length	Length  Locu  sp:STS  C)	256 s Name RAT	1.3e-40 Acc# P15589	
Protein name  Description  SULFATE SULFOHYD	ROLASE) (ARY	6051  LSULFATAS  AAID	Length  542  E C) (ASC  NT  Length	Length  Locu  Sp:STS  C)  AA  Length	256 s Name RAT Score	1.3e-40 Acc# P15589 Probability	
Protein name  Description  SULFATE SULFOHYD  ORF Name	ROLASE) (ARY	6051  LSULFATAS  AAID	Length  542  E C) (ASC  NT  Length	Length  Locu  Sp:STS  C)  AA  Length  1626  Locu	256 s Name RAT Score	Probability  9.6e-37	
Protein name  Description  SULFATE SULFOHYD  ORF Name	ROLASE) (ARY	6051  LSULFATAS  AAID	Length  542  E C) (ASC  NT  Length	Length  Locu  Sp:STS  C)  AA  Length  1626  Locu	256 s Name RAT  Score 396 s Name	Probability  9.6e-37  Acc#	

ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score	Probability
128775 f1 51		166  501	1	
Protein name		Loc	us Name	Acc#
Description				
NO-HIT				
ORF Name	NTID AAID	NT AA Length Lengt	Score	Probability
12897563 <u></u> c3 <u></u> 443	832 6054	307 924	488	1.7e-46
Protein name		Loc	us Name	Acc#
oxidoreductase, sh dehydrogenase/reduc	ort chain	pir:	72427	E72427
	case ramer			
Description				
		NT AA	Score	Probability
ORF Name	NTID AAID	NT AA Length Lengt	h Score	Probability
ORF Name			<u>Score</u>	Probability  2.1e-72
		Length Lengt	<u>h</u>	V
13.00.01 <u></u> c3 <u></u> 46.0		Length Lengt	h 383	2.1e-72
Protein name  Description	833 6055	Length Lengt  540  Loc  sp:Y	h 383  Cus Name  FCC_ECOLI	2.1e-72
Protein name  Description		Length Lengt  540  Loc  sp:Y	h 383  Cus Name  FCC_ECOLI	2.1e-72
Protein name  Description	833 6055	Length Lengt  540  Loc  sp:Y	n 383  Cus Name  FCC_ECOLI  REGION	2.1e-72 <u>Acc#</u>
Protein name  Description	833 6055	Length Lengt  540 1623  Loc  Sp:Y	n 383  Cus Name  FCC ECOLI  REGION  Score	2.1e-72
Protein name  Description  HYPOTHETICAL 54.8	KD PROTEIN IN PTA-	Length Lengt  540 1623  Loc  Sp:Y	n 383  Cus Name  FCC ECOLI  REGION  Score	2.1e-72 <u>Acc#</u>
Protein name  Description  HYPOTHETICAL 54.8  ORF Name	KD PROTEIN IN PTA-	Length Length  540  1623  Loc  Sp:Y  FOLX INTERGENIC  NT AA  Length Length  1338	n 383  Cus Name  FCC ECOLI  REGION  Score	2.1e-72 <u>Acc#</u>
Description  HYPOTHETICAL 54.8  ORF Name  13881262_t2_105	KD PROTEIN IN PTA-	Length Length  540  1623  Loc  Sp:Y  FOLX INTERGENIC  NT AA  Length Length  1338	n 383  Eus Name  FCC ECOLI  REGION  h Score	2.1e-72 Acc# Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
14331502_f3_177	835	6057	166	501		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
147.03.962 <u></u> c1334	836	6058	396	191		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA	Score	Probability
ORF Name	$\underline{\mathtt{NTID}}$	AAID	Length	Length		
14723751 <u></u> c2360	837	6059	673	2022	778	1.1e-79
Protein name		- ,		Locus	Name	Acc#
type III DNA modif (methyltransferase)		zyme		pir:F7	1810	F71810
Description				J		
200011111111111111111111111111111111111						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
14876578 <u></u> ±1 <u></u> 3.7	838	6060	296	891	310	1.2e-27
Protein name				Locus	s Name	Acc#
probable beta-glyo	cosyltransf	erase trs	C	pir:S5	1262	
Description						
			NT	AA	<del></del>	- 1 1 1 1 1 1
ORF Name	NTID	AAID	Length	Length	Score	Probability
15054691 <u></u> c1 <u></u> 3.3.0	839	6061	430	1293		
Protein name				Locu	s Name	Acc#
Description						
NO-HIT						

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
	840	6062		02		
156642_c3_469	] [010	0002	<u> </u>		: Name	Acc#
Protein name				посия	Name	110011
Description						
NO-HIT						
			NT	AA		B 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
15902 <u>f2</u> 153	841	6063	264	795	316	2.9e-28
Protein name				Locus	s Name	Acc#
				sp:YDA	ECOLI	
Description						
HYPOTHETICAL 35.6	KD PROTEIN	I IN DBPA-	INTR INTE	ERGENIC 1	REGION	
,			NT	AA		P
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name		<u>AAID</u>	Length		Score	Probability
			Length	Length	Score	Probability <u>Acc#</u>
163515 <u></u> c3 <u>466</u>			Length	Length		*
163515 <u></u> c3 <u></u> 466			Length	Length		*
163515 <u></u> c3 <u>466</u>			Length	Length		*
Protein name  Description			Length	Length	s Name	Acc#
Protein name  Description			Length	Length  Locu		*
Protein name  Description  NO-HIT	842 NTID	6064	Length 62  NT Length	Length Locu AA	s Name	Acc#
Protein name  Description  NO-HIT  ORF Name	842 NTID	6064	Length 62 NT Length	Length Locu  AA Length	s Name	Acc# Probability
Protein name Description NO-HIT ORF Name	NTID 843	6064	Length 62 NT Length	Length Locu  AA Length	Score 790 s Name	Acc# Probability  1.6e-80
Protein name  Description  NO-HIT  ORF Name  16828575_f1_43  Protein name	NTID 843	AAID 6065	NT Length	Length Locu  AA Length  1167  Locu  gp:AF0	Score 790 s Name	Acc#  Probability  1.6e-80  Acc#  AF019407

Street, St.	E.
혛	
1	diameter in
74	
April 1	1
	332
28	HIP.
Street	Haras S
1	117
Bearing	STEDEN.
1	
g)	1177
ģ	100 100 100 100 100 100 100 100 100 100
ď	114
智	11/2
Ť	
100	ud.
3	
3	
3	
3	
10 May 10	100 100 100 100
10 May 10	100 100 100 100
M. W. S. W. W.	# 11 12 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1
10 May 10	# 11 12 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1
H 41-14 II 14-14 III	
H 41-14 II 14-14 III	# 11 12 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1
10 Study 15 Study 10	
10 Study 15 Study 10	

ORF Name	NTID AAID	NT AA Score Probabil Length Length	ity
16829461_±2_112	844 6066	119 360 108 3.2e-0	6
Protein name		Locus Name Acc	<u>:#</u>
hypothetical prote	ein PH0360	pir:E71143 E711	.43
Description			
		NT AA Score Probabil	23
ORF Name	NTID AAID	Length Length Score Probabil	ity
16.8328.85 <u></u> £2 <u></u> 159	845 6067	431 1296 1723 2.3e-1	.77
Protein name		Locus Name Acc	<u>#</u>
hypothetical prote	ein	pir:JQ1020 JQ1	020
Description			
		NT AA Score Probabi	litar
ORF Name	NTID AAID	Length Length Score Probabl.	LICY
19703461c1311	846 6068	481 1446 1581 2.6e-	.62
Protein name		Locus Name Ac	<u>=#</u>
unknown	8	gp:AF048749 AF0	48749
Description			``
3	lis capsular polysa	accharide biosynthesis operon,compl	ete
sequence.			
ORF Name	NTID AAID	<u>NT AA</u> Score Probabi Length Length	lity
19725250c2343	847 6069	202 609 327 2.0e-	29
Protein name		Locus Name Ac	<u>c#</u>
		sp:YI12_METTH 027	840
Description			
HYPOTHETICAL PROT	TEIN MTH1812		

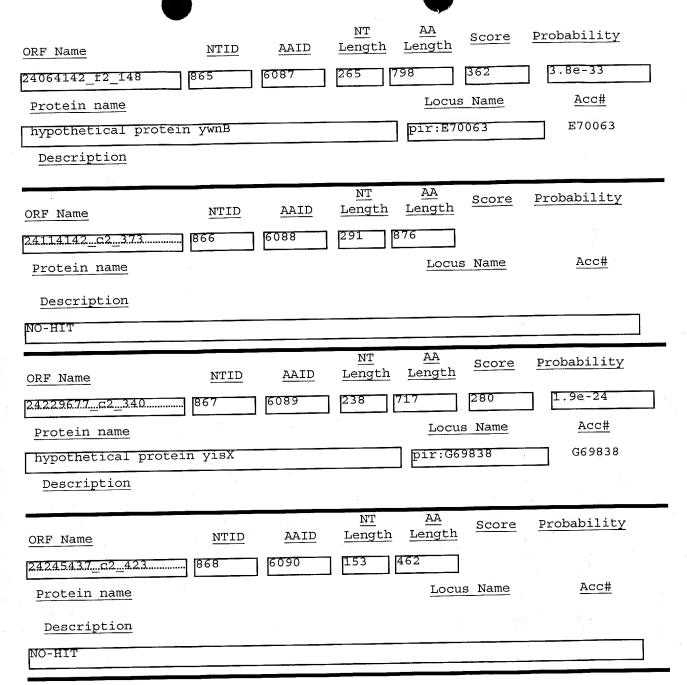
			$\underline{\text{NT}}$	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length	bcorc	11054511107
19797162_t2_119	848	6070	357	.074	1095	8.1e-111
Protein name				Locus	Name	Acc#
nucleotide sugar	epimerase			gp:AF05	9755	AF059755
Description	•					
Vibrio vulnificus	s nucleotide	sugar epi	lmerase g	gene, com	plete co	is.
		· · · · · · · · · · · · · · · · · · ·	NT	AA	Caoro	Probability
ORF Name	NTID	AAID	Length	Length	Score	PIODADITICY
1992187_f3_212	849	6071	355 1	.068	193	8.7e-15
Protein name				Locus	Name	Acc#
TumQ protein:prot	tein slr1213:	protein s	s1r1213	pir:S77	548	S77548
Description						
				7. 7.		
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 20050402_f3_253		<u>AAID</u>	Length		Score	Probability  1.8e-07
			Length	Length		
20050402 <u></u> £3 <u></u> 253	850		Length	Length	129 s Name	1.8e-07
20050402_f3_253 Protein name	850		Length	Length  192  Locus	129 s Name	1.8e-07 <u>Acc#</u>
Protein name  phosphopyruvate	850		Length	Length  192  Locus  pir:C79	129 s Name	1.8e-07 <u>Acc#</u>
Protein name  phosphopyruvate	850		Length	Length  192  Locus	129 s Name	1.8e-07 <u>Acc#</u>
Protein name  phosphopyruvate  Description	hydratase  NTID	6072	Length  NT Length	Length Locus pir:C75	129 3 Name 5251	1.8e-07 Acc# C75251
Protein name  Phosphopyruvate  Description  ORF Name	hydratase  NTID	AAID	Length  163  NT Length	Length Locus pir:C75  AA Length	129 S Name 5251 Score	1.8e-07 <u>Acc#</u> C75251  Probability
Protein name  Protein name  phosphopyruvate  Description  ORF Name  20087751_t3_199	hydratase  NTID	AAID 6073	Length  163  NT Length  319	Length Locus pir:C75  AA Length	129 S Name S251 Score 1657 S Name	1.8e-07 Acc# C75251  Probability 2.3e-170
Protein name  phosphopyruvate  Description  ORF Name  20087751_t3_199	hydratase  NTID	AAID 6073	Length  163  NT Length  319	Length Locus Locus AA Length Locus	129 S Name S251 Score 1657 S Name	1.8e-07 Acc# C75251  Probability  2.3e-170 Acc#
Protein name  phosphopyruvate  Description  ORF Name  20087751_t3_199  Protein name  putative UDP-Glc	NTID  NAC: undecapre	AAID 6073 enylphosp	Length  NT Length  319	Length Locus  Locus  AA Length  Locus  gp:AF0	129 Sore 1657 Name 48749	1.8e-07 Acc# C75251  Probability  2.3e-170 Acc# AF048749

ORF Name	NTID 2	AAID I	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2037502_£1_17	852 60'	74 2	256 7	71	245	9.6e-21
Protein name	<u> </u>			Locus	Name	Acc#
conserved hypothet	ical protein			pir:D72	2320	D72320
Description						
ORF Name	NTID	AAID I	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
20734625 £3208	853 60	75 2	223 6	72	225	1.3e-18
Protein name				Locus	s Name	Acc#
hypothetical prote	in	1		gp:SSU	18930	Y18930
Description				L	0	
Sulfolobus solfata	iricus 281 kb	genomic	DNA fra	gment,	strain P2	2.
			NT	AA		Probability
ORF Name	NTID	AAID	Length	Length	Score	PIODADITICY
20754427 <u></u> f3 <u></u> 198	854 60	76	347 1	044	1691	5.7e-174
Protein name				Locu	s Name	Acc#
UDP-glucose-4-epin	nerase/dTDP-gl	ucose-4,	, 6	gp:AF0	48749	AF048749
Description						*
Bacteroides fragi	lis capsular r	olysacci	haride b	iosvnth	esis ope	ron, complete
sequence.		- 8	·			
		8	NT	AA	Score	Probability
sequence.  ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
sequence.	NTID	AAID	<u>NT</u> Length	AA Length	Score 307	Probability  2.6e-27
ORF Name  211510_f1_57  Protein name	NTID 855	AAID	<u>NT</u> Length	AA Length 029	Score 307 as Name	Probability  2.6e-27  Acc#
ORF Name  211510t157  Protein name  activator protein	NTID 855	AAID	<u>NT</u> Length	AA Length	Score 307 as Name	Probability  2.6e-27
ORF Name  211510_f1_57  Protein name	NTID 855	AAID 077	NT Length	AA Length .029 Locu	Score 307 us Name 047527	Probability  2.6e-27  Acc#  AF047527

St. of	
ik.	24 24 24
, mil	M
:::1	#37
ť,	axi.
	14 THE
# Thresh and	115
15,23	SHE SHE
<b>27</b>	
明の世	Harris Harris
stra	12774
Buch	HATTE OF
	353
100	
13046	
ř	77

ORF Name	NTID AAID	NT Length	AA Length Score	Probability
21640887_f2_117	856 6078	362	1089 157	1.7e-08
Protein name			Locus Name	Acc#
hypothetical prot	ein 7.17	A .	pir:D47677	
Description				
		NT	AA Grava	D 1 -1 -2 1 -2 hora
ORF Name	NTID AAII		Length Score	Probability
21681552 <u></u> c3 <u></u> 442	857 6079	466	1401 793	8.2e-79
Protein name			Locus Name	Acc#
thiophene and fur	ran oxidation prote	ein	pir:C70375	C70375
Description				
		NT	AA Ggoro	Probability
ORF Name	NTID AAII		Length Score	Probability
22536527 <u>f3</u> _190	858 6080	215	648 1136	3.7e-115
Protein name			Locus Name	Acc#
putative methyl	transferase		gp:AF048749	AF048749
Description			*	
	ilis capsular poly	saccharide	biosynthesis op	eron, complete
sequence.				
ORF Name	NTID AAI	<u>NT</u> D Length	AA Score	Probability
22540937_£1_4	859 6081	522	1569 304	1.1e-45
Protein name		<u> </u>	Locus Name	Acc#
			sp:STS_HUMAN	P08842
Description				
SULFATE SULFOHYI	DROLASE) (ARYLSULFA	ATASE C) (A	SC)	

ORF Name	NTID AAID	NT AA Length Length	Score Pro	bability
22860128_f3_255 860	6082	83 252	64 0	.031
Protein name		Locus	s Name	Acc#
		sp:SPR	C_XENLA	P36378
Description				
(OSTEONECTIN) (ON) (BA	SEMENT MEMBRANE	PROTEIN BM-40)		
		NT AA	Score Pro	bability
ORF Name	NTID AAID	Length Length		70.00
23472533_f3_254 86:	6083	124 375		3.7e-08
Protein name			s Name	Acc#
phosphopyruvate hydrat	ase	pir:C7	5251	C75251
Description				
		NT AA	Score Pro	obability
	NTID AAID	Length Length		
2347.9066c3434	2 6084	183 552	, , , , , , , , , , , , , , , , , , ,	7 mmH
Protein name		Locu	s Name	Acc#
Description				.00
NO-HIT			81	
		NT AA	Score Pro	obability
ORF Name	NTID AAID	Length Length	×	
23647758 <u>c2</u> 399 86	3 6085	81 246		
Protein name		Locu	s Name	Acc#
Description				
NO-HIT			×	
1		<u>NT</u> <u>AA</u>	Score Pr	obability
ORF Name	NTID AAID	Length Length		
23.70.9625 <u></u> c343386	6086	321 966		3.2e-61
Protein name			ıs Name	Acc#
dolichol-phosphate mar	inosyltransferase	e pir:G7	70463	G70463
Description				



ODE Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name [24257187 f2_132		91		.236	1104	9.0e-112
				Locus	Name	Acc#
Protein name						
putative carboxybio	otin decarbox	ylase s	ubunit	gp:MRU	37980	U87980
Description						
Malonomonas rubra decarboxylase gene madH, madK, madF, m	cluster (mad)	z, madz,	madG, n	nadB, mad	dA, madE,	alonate madC, madD,
		77.7	NT Length	<u>AA</u> Length	Score	Probability
ORF Name	NTID	AAID	Length			
24401507_c1_299	870 60	092	510	1533	2702	4.2e-281
Protein name	*			Locu	s Name	Acc#
unknown				gp:AF0	48749	AF048749
Description						
Bacteroides fragil sequence.	is capsular p	polysac	charide l	biosynth	esis ope	ron, complete
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>	<u>50020</u>	*
24473192 <u></u> c2 <u></u> 372	871 6	093	642	1929	110	0.0037
Protein name				Locu	s Name	Acc#
				sp:Y0E	W_MYCLE	Q49757
Description						· ·
HYPOTHETICAL 31.1	KD PROTEIN B	1937_F2	_39		18.1	×-

Protein name  Locus Name  Acc#  putative hemolysin  Description  Streptococcus mutans YtqB (ytqB) gene, partial cds; ABC transporter(abcX), putative permease (perM), putative hemolysin (hlyX), pyruvate-formate lyase activating enzyme (pflC), D-alanine-D-alanylcarrier protein ligase (dltA), integral membrane protein (dltB), D-alanyl carrier protein (dltC), extramembranal protein (dltD), andputative exopolyphosphatase (ppxl) genes,			NT AA Score Probability
Protein name	ORF Name		
Description   Sp:AF051356   AF051356	24473817_£2_106	872 6094	431 1296 457 3.3e-43
Description  Streptococcus mutans YtqB (ytqB) gene, partial cds; ABC transporter(abcX), putative permease (perM), putative hemolysin (hlyX), pyruvate-formate lyase activating enzyme (pfIC), D-alanine-D-alanylcarrier protein ligase (dltA), integral membrane protein (dltB), D-alanyl carrier protein (dltC), extramembranal protein (dltB), andputative exopolyphosphatase (ppxl) genes, extramembranal protein (dltD), andputative exopolyphosphatase (ppxl) genes, extramembranal protein figure probability  Protein name	Protein name		Locus Name Acc#
Streptococcus mutans YtqB (ytqB) gene, partial cds; ABC transporter (abcX), putative permease (perM), putative hemolysin (hlyX), pyruvate-formate lyase activating enzyme (pfIC), D-alanine-D-alanylcarrier protein ligase (dltA), integral membrane protein (dltB), D-alanyl carrier protein (dltC), extramembranal protein (dltD), andputative exopolyphosphatase (ppxl) genes, extramembranal protein figures,	putative hemolysi	n	gp:AF051356 AF051356
putative permease (perM), putative hemolysin (hlyx), pyruvate-formate lyase activating enzyme (pflC), D-alanine-D-alanylcarrier protein ligase (dltA), integral membrane protein (dltB), D-alanyl carrier protein (dltC), extramembranal protein (dltD), and putative exopolyphosphatase (ppx1) genes,  ORF Name  NTID  AAID  Length  Length  Length  Length  Protein name  Locus Name  Acc#  AF145800  Description  Porphyromonas gingivalis strain W50 immunogenic 75 kDa protein PG4gene, complete cds.  ORF Name  NTID  AAID  Length  Length  Length  Length  Length  Length  Score  Probability  AAA  ORF Name  NTID  AAID  Length  Length  Length  Length  Length  Length  Score  Probability  Description  Bacteroides iragilis capsular polysaccharide biosynthesis operon, complete sequence.  ORF Name  NTID  AAID  Length  Length  Length  Length  Locus Name  Acc#  ACC#  Sp:AF048749  AF048749  AF048749  Description  Bacteroides iragilis capsular polysaccharide biosynthesis operon, complete sequence.  ORF Name  NTID  AAID  Length  Length  Length  Length  Score  Probability  AACC#  Sp:RIBS_ECOLI  P24199  Description			
ORF Name         NTID         AAID         Length         Length         Score         PROBABILITY           244953337_c2_345         873         6095         499         1500         119         0.00014           Protein name         Locus Name         Acc#           1mmunogenic 75 kba protein PG4         gp:AF145800         AF145800           Description         Porphyromonas gingivalis strain W50 immunogenic 75 kba protein PG4gene, complete cds.         NTID         AAID         NTID         AAID         Length         Length         PC4gene, complete cds.           ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         Locus Name         Acc#           Unknown         gp:AF048749         AF048749         AF048749           Description         Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.         NTID         AAID         Length         Length         Score         Probability           ORF Name         NTID         AAID         Length         Length         Score         Probability           245447.763_cl_275	putative permease activating enzyme	(perM), putative hem (pflC), D-alanine-D-a protein (dltB), D-alan	olysin (hlyX),pyruvate-formate fyase alanylcarrier protein ligase (dltA), nyl carrier protein (dltC),
Protein name	ORF Name	NTID AAID	Score Probability
Immunogenic 75 kDa protein PG4	24495337 <u></u> c2 <u></u> 345	873 6095	1500 119 0.00014
Description	Protein name		Locus Name Acc#
Porphyromonas gingivalis strain W50 immunogenic 75 kDa protein PG4gene, complete cds.    NT	immunogenic 75 kl	a protein PG4	gp:AF145800 AF145800
ORF Name  NTID  AAID  Length  Length  Score  Probability  24631305_#2_116	Description		
ORF Name         NTID         AAID         Length         Length           24631305_#2_116		ngivalis strain W50 i	NT AA Score Probability
Protein name         Locus Name         Acc#           unknown         gp:AF048749         AF048749           Description         Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.           ORF Name         NTID         AAID         Length         Score         Probability           24647763_c1_275         875         6097         498         1497         685         2.3e-67           Protein name         Locus Name         Acc#           Sp:RIBB_ECOLI         P24199           Description	ORF Name	NTID AAID	Length Length
unknown         gp:AF048749         AF048749           Description         Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.           ORF Name         NTID         AAID         Length         Score         Probability           24647763_c1_275	246313.05 <u></u> ±2 <u>.</u> 116	874 6096	183 552 575 1.0e-55
Description  Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.  ORF Name  NTID  AAID  Length  Length  Length  Score  Probability  2.3e-67  Protein name  Locus Name  Acc#  Sp:RIBB_ECOLT  P24199  Description	Protein name		Locus Name Acc#
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.    NT AA Score Probability	unknown		gp:AF048749 AF048749
ORF Name         NTID         AAID         NT Length         AA Length         Score         Probability           24647763cl275	Description		
ORF Name         NTID         AAID         Length         Length         Score         Probability           2464.77.63_c1_275		ilis capsular polysac	charide biosynthesis operon,complete
ORF Name         NTID         AAID         Length         Length           24647763c1275			NT AA Ggara Probability
Protein name         Locus Name         Acc#           sp:RIBB_ECOLI         P24199   Description	ORF Name	NTID AAID	Length Length Score Flooability
pescription p24199	24647763 <u>c1</u> 275	<u>875</u> 6097	498 [1497 [685 2.3e-67
Description Description	Protein name	ī	Locus Name Acc#
			sp:RIBB_ECOLI P24199
	Description		
		BUTANONE 4-PHOSPHATE	SYNTHASE (DHBP SYNTHASE)

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24651515_£2_91	876	6098	108	Jogus	s Name	Acc#
Protein name				посия	5 Name	1100 m
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24666000 <u></u> c3 <u></u> 437	877	6099	306	921	150	4.1e-16
Protein name				Locu	s Name	Acc#
probable uridine p	hosphoryla	se APE210	)5	pir:D7	2516	D72516
Description	-	4				1.
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24804663 <u>f3</u> 195	878	6100	386	1161		
Protein name				Locu	s Name	Acc#
Description		9				· · ·
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
25972937 <u></u> c3 <u></u> 459	879	6101	69	210		
Protein name				Locu	ıs Name	Acc#
Description				*		
NO-HIT						-
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
26212777 <u>c3</u> 450	880	6102	300	903	215	1.4e-17
Protein name				Loci	ıs Name	Acc#
hypothetical prot	ein sll167	1.		pir:S	74655	S74655
Description						

ORF Name	NTID AAID	NT AA Length Length	re <u>Probability</u>
26261313_f1_58	881 6103	63 192	
Protein name		Locus Nam	<u> Acc#</u>
Description			
NO-HIT			
		NT AA Coo	Dechability
ORF Name	NTID AAID	Length Length	
26444687 <u></u> c2341	882 6104	1305 1588	4.6e-163
Protein name		Locus Nam	e Acc#
		sp:ENO_STAA	O69174
Description			
GLYCERATE HYDRO-L	YASE) (LAMININ BINDI	NG PROTEIN)	
		NT AA G	
ORF Name	NTID AAID	Length Length	
26594206 <u></u> £1 <u></u> 42	883 6105	180 543 896	9.9e-90
Protein name		Locus Nan	ne Acc#
putative hypoxant	hine guanine	gp:AF048749	AF048749
Description			
Bacteroides fragi sequence.	lis capsular polysac	ccharide biosynthesis	operon, complete
		NT AA G	
ORF Name	NTID AAID	Length Length	ore Probability
26594686 <u></u> c2 <u></u> 3.46	884 6106	158 477	
Protein name		Locus Na	me Acc#
Description			•
			¥

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
26595337_±3_180	885	6107	175	528		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
				7.7		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
26664012 <u></u> c2 <u></u> 404	886	6108	355	1068	112	0.0032
Protein name				Locus	s Name	Acc#
gamma response I p	rotein			gp:ATH	131708	AJ131708
Description						
Arabidopsis thalia	ana gr I ge	ne, exons	1-3.			a 1
			NT	AA		Drobobility
ORF Name	NTID	AAID	Length	Length	Score	Probability
26761057 <u></u> c2 <u></u> 361	887	6109	1017	3054	627	1.0e-119
Protein name				Locu	s Name	Acc#
restriction endon	ıclease			gp:AF0	60119	AF060119
Description	1 8					
Pasteurella haemo (res) genes, compl	lytica meth ete cds.	yltranste	rase (mo	od) and r	estricti	onendonuclease
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>		Score	<u>Probability</u>
26836680 <u></u> 12 <u></u> 109	888	6110	416	1251	1189	8.9e-121
Protein name				Locu	s Name	Acc#
immunoreactive 47	kD antiger	PG120		gp:AF1	44640	AF144640
Description		*				
Porphyromonas gin complete cds.	givalis stī	ain W50 i	Lmmunorea	active 47	kD anti	genPG120 gene,

	MULT	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name	NTID			410	111	2.7e-06	
275125_£2_111	889	6111			<u></u>	Acc#	
<u>Protein name</u>				Locus			
hypothetical prot	ein Rv2333c			pir:F70	705	F70705	
Description							
			NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length	30016	11000011101	
2812827 <u></u> £3 <u></u> 207	890	6112	270	313	298	2.3e-26	
Protein name				Locus	Name	Acc#	
				sp:YFIF	_HAEIN	P44552	
Description						i .	
HYPOTHETICAL PROT	TEIN HI0175						
			NT	AA		n 1 -1-2124	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 29336040_f1_36		<u>AAID</u>	Length		Score	Probability 7.9e-07	]
			Length	Length	*		]
29336.040 <u></u> f1 <u></u> 36	891	6113	Length 420	Length	144 s Name	7.9e-07	]
29336040f136 Protein name	891	6113	Length 420	Length  1263  Locus	144 s Name	7.9e-07 <u>Acc#</u>	]
Protein name  NADH dehydrogenas	891	6113	Length 420	Length  1263  Locus  pir:T1	144 s Name	7.9e-07 <u>Acc#</u>	]
Protein name  NADH dehydrogenas  Description	891	6113	Length 420	Length  1263  Locus	144 s Name	7.9e-07 <u>Acc#</u>	]
Protein name  NADH dehydrogenas  Description  ORF Name	se (ubiquino	one), chai	Length 420  n 2  NT Length	Length  1263  Locus  pir:T1:	144 s Name 1319	7.9e-07 Acc# T11319	]
Protein name  NADH dehydrogenas  Description  ORF Name	se (ubiquino	one), chai	Length 420  n 2  NT Length	Length  Locus  pir:T1  AA  Length	144 S Name 1319 Score	7.9e-07 Acc# T11319 Probability	]
Protein name  NADH dehydrogenas  Description  ORF Name	se (ubiquino	one), chai	Length 420  n 2  NT Length	Length  Locus  Locus  AA  Length  1191  Locu	144 S Name 1319 Score 369 S Name	7.9e-07 Acc# T11319  Probability 6.9e-34	]
Protein name  NADH dehydrogenas  Description  ORF Name  29703165_t2_118	se (ubiquino	one), chai	Length 420  n 2  NT Length	Length  Locus  Locus  AA  Length  1191  Locu	144 S Name 1319 Score	7.9e-07 Acc# T11319  Probability 6.9e-34 Acc#	]
Protein name  NADH dehydrogenas  Description  ORF Name	se (ubiquino	one), chai	Length 420  n 2  NT Length	Length  Locus  Locus  AA  Length  1191  Locu	144 S Name 1319 Score 369 S Name	7.9e-07 Acc# T11319  Probability 6.9e-34 Acc#	]

$\frac{\text{NT}}{\text{ORF Name}}  \frac{\text{AA}}{\text{NTID}}  \frac{\text{AAID}}{\text{Length}}  \frac{\text{Ength}}{\text{Length}}  \frac{\text{Score}}{\text{Probability}}$	
30084688_f2_127	
Description	
NO-HIT	
ORF Name NTID AAID Length Length Score Probability	
3.023452 <u></u> c3 <u></u> 522	
Protein name Locus Name Acc#	
alkaline phosphatase gp:SSPPHOA2 Z48801	
Description	
Synechococcus PCC7942 phoV gene for alkaline phosphatase.	
NT AA Score Probability	*
ORF Name NTID AAID Length Length  [31220027 G2 247   1895   16117   182   1549   204   6.9e-15	
31330037d2347	
Protein name Locus Name Acc#	
DNA polymerase III, alpha subunit pir:C72360 C72360	
Description	
ORF Name NTID AAID Length Length Score Probability	
3240675f3_182896   6118   135   408   115   7.2e-07	
Protein name Locus Name Acc#	
protein-export membrane protein pir:E71837 E71837	
Description	
NT AA Grave Probability	
ORF Name NTID AAID Length Length Score Probability	
33301250c2_380	
Protein name Locus Name Acc#	
Description	
NO-HIT	

ORF Name	NTID AAID	NT AA Score Probability
33397811_±1_45	898 6120	898 2697 115 0.00042
Protein name		Locus Name Acc#
histidine kinase s	ensor protein	pir:D70328 D70328
Description		
		NT AA Grave Drobability
ORF Name	NTID AAID	Length Length Score Probability
33489041_f2_110	899 6121	260 783 94 0.045
Protein name		Locus Name Acc#
		sp:TPMN_XENLA Q01174
Description	•	
TROPOMYOSIN ALPHA	CHAIN, NON MUSCLE	
		NT AA Score Probability
ORF Name	NTID AAID	Length Length ———
33881887 <u></u> £3 <u></u> 192	900 6122	83 252 69 0.042
Protein name		Locus Name Acc#
		sp:YA49_HAEIN
Description		
HYPOTHETICAL PROT	EIN HI1049	
ORF Name	NTID AAID	NT AA Score Probability Length Length
34189385f138	901 6123	311 936 514 3.0e-49
Protein name		Locus Name Acc#
		gp:BCY11138 Y11138
Description	•	
B.cereus DNA for		

	*		$\underline{\mathrm{NT}}$	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length	50025		
34407193_f1_47	902	6124	144 4	35			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA	G	Drobobility	_
ORF Name	NTID	AAID	Length	Length	Score	Probability	
34412513£3194	903	6125	303	912	193	1.4e-14	]
Protein name				Locus	Name	Acc#	
glycosyl transfe	rase PAB0772			pir:B75	096	B75096	
Description	,						
			2777	7) 7)	<u> </u>		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
34640915f153	904	6126	63	192			
Protein name				Locus	s Name	Acc#	
Description							
NO-HIT			*		*		
/ 1							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
34650341 £3 230	905	6127	79	240			
Protein name		<u> </u>	<u></u>	Locu	s Name	Acc#	
Description							
	*					-	
NO-HIT							
			NT	<u>AA</u>	Score	Probability	
ORF Name	NTID	AAID	Length	<u>Length</u>		- F1	_
34652167 <u></u> £2 <u></u> 108	906	6128	375	1128	530	6.0e-51	
Protein name		10			s Name	Acc#	
pyridoxal phosph	nate biosynthe	etic prot	ein PdxA	pir:H7	0373	H70373	
Description							

ORE Name NTID AAID Length Length
ORT Name
Jagua Nama Acc#
Protein name Locus Name
Description
NO-HIT
NT AA Score Probability
ORF Name NTID AAID Length Length
3.3e-18 3.3e-18
Protein name Locus Name Acc#
hypothetical protein pir:H75473 H75473
Description
NT AA Goorg Brobability
ORF Name NTID AAID Length Length Probability
3944005t139909
Protein name Locus Name Acc#
putative UDP-glucose dehydrogenase gp:AF159428 AF159428
Description
Burkholderia pseudomallei putative UDP-glucose dehydrogenase (udg), putative
ADP-heptose synthase (waaE), and putativeADP-glycero-mannoheptose epimerase (gmhD) genes, complete cds.
$rac{ ext{NT}}{ ext{ORF Name}} rac{ ext{AA}}{ ext{NTID}} rac{ ext{AAID}}{ ext{Length}} rac{ ext{Eength}}{ ext{Length}}$
3988318_c2_384
7.00
Procein name
putative TonB-dependent outer membrane gp:AF048749 AF048749 receptor
Description
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete
sequence.

ORF Name	NT AA Score Probability NTID AAID Length Length
3990900_£3_193	911 6133 392 1179 272 1.3e-31
Protein name	Locus Name Acc#
probable galacto	yltransferase trsD pir:S51263
Description	
	NT AA Score Probability
ORF Name	NTID AAID Length Length
4065757 <u></u> c1 <u></u> 259	
Protein name	Locus Name Acc#
Description	
NO-HIT	
	NT AA Score Probability
ORF Name	NTID AAID Length Length
4094512 <u>c3</u> 436	913 6135 344 1035 1054 1.8e-106
Protein name	Locus Name Acc#
	sp:YQFA_BACSU P54466
Description	
HYPOTHETICAL 35.	6 KD PROTEIN IN RPSU-PHOH INTEREGENIC REGION
ORF Name	NTID AAID Length Length Score Probability
4116680 <u></u> ±3 <u></u> 175	914 6136 245 738
Protein name	Locus Name Acc#
Description	
NO-HIT	
	NT AA Score Probability
ORF Name	NTID AAID Length Length
4119677 <u></u> £2 <u></u> 125	
Protein name	Locus Name Acc#
hypothetical pro	tein jhp1456 pir:C71806 C71806
Description	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4147280 C2 374		6138		204		
	الستاا		L_		: Name	Acc#
Protein name				10000		<u></u>
Description						
NO-HIT						
X.			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length		
425010 <u></u> £3 <u></u> 196	. 917	6139	385	1158	642	8.2e-63
Protein name				Locus	Name	Acc#
WbpU				gp:AF0	35937	AF035937
Description						
Pseudomonas aerugi Ihf-Beta, Wzz (wzz) for O-antigen biosy	, and Wzx	(wzx) gen	es, compi	psA) gene lete cds;	e, parti andwbp	alcds; gene cluster
ior o anergen zree,		7			(#	
for a distinguished,		T	NT	AA	Score	Probability
ORF Name	NTID	AAID	NT Length	Length	*1	*
	NTID		NT Length	<del></del>	Score	1.6e-06
ORF Name	NTID	AAID	NT Length	Length  1422  Locus	143 s Name	1.6e-06 <u>Acc#</u>
ORF Name 427215_f3_162	NTID	AAID	NT Length	Length	143 s Name	1.6e-06
ORF Name  427215_f3_162  Protein name  unknown  Description	<u>NTID</u> 918	<u>AAID</u> 6140	NT Length 473	Length  Locus  Locus  gp: U96	143 S Name	1.6e-06 <u>Acc#</u> U96771
ORF Name 427215_f3_162 Protein name unknown	NTID 918	AAID 6140 polygala	NT Length 473	Length  Locus  Locus  gp: U96	143 S Name	1.6e-06 <u>Acc#</u> U96771
ORF Name  427215f3162  Protein name  unknown  Description  Prevotella bryanti	NTID 918 putative omplete cds	AAID 6140 polygala and unk	NT Length  473	Length  Locus  gp:U96  e,B-1,4-6  s.  AA	143 S Name	1.6e-06 <u>Acc#</u> U96771
ORF Name  427215f3162  Protein name  unknown  Description  Prevotella bryanti	NTID  i putative omplete cds	AAID 6140 polygala ; and unk	NT Length  473  cturonas nowngene  NT Length	Length  Locus  Locus  gp:U96  e,B-1,4-6  s.  AA  Length	143 S Name 771 endogluc	1.6e-06 Acc# U96771 anase, and
ORF Name  427215f3162  Protein name  unknown  Description  Prevotella bryantimannanase genes, co	NTID  i putative omplete cds  NTID	AAID 6140 polygala and unk	NT Length  473  cturonas nowngene  NT Length	Length Locus Locus gp:U96 e,B-1,4-6 s.  AA Length 2868	143 s Name 771 endogluc Score	1.6e-06  Acc#  U96771  anase, and  Probability
ORF Name  427215_f3_162  Protein name  unknown  Description  Prevotella bryantimannanase genes, co	NTID  i putative omplete cds  NTID	AAID 6140 polygala ; and unk	NT Length  473  cturonas nowngene  NT Length	Length Locus Locus gp:U96 e,B-1,4-6 s.  AA Length 2868	143 S Name 771 endogluc	1.6e-06 Acc# U96771 anase, and
ORF Name  427215_f3_162  Protein name  unknown  Description  Prevotella bryantimannanase genes, co	NTID  i putative omplete cds  NTID	AAID 6140 polygala ; and unk	NT Length  473  cturonas nowngene  NT Length	Length Locus Locus gp:U96 e,B-1,4-6 s.  AA Length 2868	143 s Name 771 endogluc Score	1.6e-06  Acc#  U96771  anase, and  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4331300_c2_390		6142	82	249	103	1.1e-05
Protein name				Locus	s Name	Acc#
hypothetical prote	ein			gp:SSU	18930	Y18930
Description						
Sulfolobus solfata	aricus 281 k	b genomi	c DNA fr	agment,	strain Pi	2.
			NT	AA	G	Dechahilitu
ORF Name	NTID	AAID	Length	Length	Score	Probability
4409462_c2_403	921	6143	506	1521	631	1.2e-61
Protein name				Locu	s Name	Acc#
conserved hypothe	tical protei	n aq_136	5	pir:F7	0418	F70418
Description						
				7.7		
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
46.941.52 <u></u> £2 <u></u> 10.7	922	6144	717	2154	122	0.00078
Protein name				Locu	s Name	Acc#
putative peptidyl	-prolyl cis-	trans is	omerase	gp:ASA	J2316	AJ002316
Description						
Acinetobacter sp.	ADP1 alkR &	alkM ge	enes, ORI	rl & ORF4		
			NT	AA		
ORF Name	NTID	AAID	Length		Score	Probability
4725257 <u>f2104</u>	923	6145	427	1284	89	0.011
Protein name				Loci	ıs Name	Acc#
membrane protein	<sup>34</sup> .			gp:PP(	JY18245	Y18245
Description						
Pseudomonas putio	ia todX, tod!	F, todC1	, todC2,	todB, to	dA, todl	),todE, todG,
todI, todH, todS,	todT genes.	4				

	_
ORF Name	$rac{ ext{NT}}{ ext{NTID}}$ AAID Length Length
4804632_c3_476	924 6146 218 657 1119 2.3e-113
Protein name	Locus Name Acc#
unknown	gp:AF048749 AF048749
Description	
Bacteroides frag	lis capsular polysaccharide biosynthesis operon,complete
	NT AA Grane Probability
ORF Name	NTID AAID Length Length Score Probability
5110700_f1_35	925 6147 485 1458 780 1.9e-77
Protein name	Locus Name Acc#
O-antigen repeat	unit transporter Wzx gp:AF172324 AF172324
Description	
transporter Wzx (wbnB), WbnC (wbn	GalF (galF) gene, partial cds; O-antigen repeatunit wzx), WbnA (wbnA), O-antigen polymerase Wzy(wzy), WbnB C), WbnD (wbnD), WbnE (wbnE), UDP-Glc-4-epimerase GalE gluconate dehydrogenaseGnd (gnd), UDP-Glc-6-dehydrogenase nF (wbnF)genes, complete cds; and chain length determinant
ORF Name	NT AA Score Probability NTID AAID Length Length
52.70252 <u></u> f2 <u></u> 124	926 6148 192 579 335 3.4e-40
Protein name	Locus Name Acc#
	gp:AB017508 AB017508
Description	
Bacillus halodur	ans C-125 genomic DNA, 32 kb fragment, completecds.
	NT AA Score Probability
ORF Name	NTID AAID Length Length
5313.7.75 <u></u> ±1 <u></u> 3.4	
Protein name	Locus Name Acc#
unknown	gp:AF048749 AF048749
Description	
	gilis capsular polysaccharide biosynthesis operon, complete
sequence.	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
553212 f3 241	928	6150	654	965	485	6.4e-45
Protein name			L	Locus	s Name	Acc#
sensory transducti	on histidir	ie kinase		pir:S7	5130	S75130
slr2098:protein slr	2098:protei	n slr209	8			
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
657956_c1_258	929	6151	210	633		
Protein name				Locu	s Name	Acc#
Description						
NO-HIT						
	*		NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length		
6671887 <u>f3</u> 238	930	6152	1294	3885	384	6.5e-33
Protein name					s Name	Acc#
putative alpha-gl	ucosidase			gp:AAC	252161	AJ252161
Description						- Land on
Alicyclobacillus (malEFGR genes, cd	acidocaldar aA gene and	ius malt IglcA ge	ose/malto ne).	dextrine	transpo	rtgene region
			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length		602	1.4e-58
6.687.552 <u></u> £1 <u></u> 19	931	6153	349	1050	j	
Protein name					ıs Name	Acc# P36979
				sp:YF	GB_ECOLI	F309/9
Description	9				NICION	
HYPOTHETICAL 43.1	L KD PROTEII	IN NDK-	GCPE INT	ERGENIC 1	KEG TON	

ORF Name	NTID AAID	NT AA Length Length Score	Probability
6767537_f1_13  Protein name	]	Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Length Length Score	Probability
6854762c3470 Protein name	933 6155	Locus Name	Acc#
Description	···		
NO TITE			
ORF Name	NTID AAID	NT AA Score Length Length	Probability
7225916f183 Protein name	934 6156	Locus Name	Acc#
Description		-	,
NO-HIT			
ORF Name	NTID AAID	NT AA Length Length Score	Probability
7.96.875f2158	935 6157	60 183	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
961067_c3_479 Protein name	936 6158	Locus Name	Acc#
Description			
NO-HIT			

ORF Name NTID AAID Length Length  ORF Name NTID NAME NAME NTID NAME NAME NAME NAME NAME NAME NAME NAME
9813 c3 473 937 6159 65 198
Protein name Locus Name Acc#
Description
NO-HIT
NT AA Score Probability ORF Name NTID AAID Length Length
9921927_c1_288938 6160 419 1260 568 5.7e-55
Protein name Locus Name Acc#
conserved hypothetical protein ykgB pir:D69856 D69856
Description
NT AA Guun Duchahilitu
ORF Name NTID AAID Length Length  ORF Name
10009632
Protein name Locus Name Acc#
unknown gp: U96771 U96771
Description
Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknowngenes.
NT AA Score Probability
ORF Name NTID AAID Length Length Probability
10753760_f2_32
Protein name Locus Name Acc#
Protein name Locus Name Acc#
IgA Fc receptor-like protein A428L pir:T17931 T17931

Stanfe Stanfe	150 150 150 150 150 150 150 150 150 150
Ì,	ing.
Taraff.	Fi
	455
834	Dist.
Tariff.	7
STREET,	
Sec. 3	22
1	327 327
#	
1	11
	in:
30	1:351
Store.	100 mg
	pp.
161	500
17 30 45	74
No.	nati

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
12787768_f3_65  Protein name	941	5163	0 /		Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
13953388 <u></u> c1 <u></u> 94	942	6164	296	391.	133	2.2e-06	
Protein name					Name	Acc#	
				sp:VIR	YEREN	P13225	
Description							
	THE TONIST THE	アスギカイ カーワ	TVATOR				
VIRULENCE REGULON	TRANSCRIPTI	ONAL ACT	IVATOR				
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability	
	NTID		<u>NT</u> Length		Score	Probability	
ORF Name	NTID	AAID	<u>NT</u> Length	Length 993 Locu	212 s Name	1.5e-18 <u>Acc#</u>	_
ORF Name 14511086cl86	NTID 943	AAID	<u>NT</u> Length	Length	212 s Name	1.5e-18	
ORF Name  14511086cl86	NTID 943	AAID	<u>NT</u> Length	Length 993 Locu	212 s Name	1.5e-18 <u>Acc#</u>	
ORF Name  14511086cl86  Protein name  hypothetical protein	NTID 943	AAID	<u>NT</u> Length	Length 993 Locu	212 s Name	1.5e-18 <u>Acc#</u>	
ORF Name  14511086cl86  Protein name  hypothetical protein description	NTID 943 ein F14F9.5	<u>AAID</u> 6165	NT Length	Length 993  Locu pir:T3	212 s Name 3774	1.5e-18 Acc# T33774	
ORF Name  14511086cl86  Protein name  hypothetical protection  Description  ORF Name	NTID 943 ein F14F9.5	AAID 6165 AAID	NT Length  330  NT Length	Length  993  Locu  pir:T3  AA  Length  1296  Locu	212 s Name 3774 Score 1723 s Name	1.5e-18  Acc#  T33774  Probability  2.3e-177  Acc#	
ORF Name  14511086cl86  Protein name  hypothetical protein description  ORF Name  16832885t368	NTID 943 ein F14F9.5  NTID 944	AAID 6165 AAID	NT Length  330  NT Length	Length  993  Locu  pir:T3  AA  Length  1296	212 s Name 3774 Score 1723 s Name	1.5e-18  Acc#  T33774  Probability  2.3e-177	

	*****************	3 7 TT	NT	<u>AA</u> Length	Score	Probability
ORF Name	NTID	AAID	Length		12.4	] [0.031
22860128_f1_11	945	6167	83 2	52	64	
Protein name				Locus	Name	Acc#
				sp:SPRC	_XENLA	P36378
Description						
(OSTEONECTIN) (ON)	(BASEMENT	MEMBRANE	PROTEIN	BM-40)	T.	
			NT	AA		
ORF Name	NTID	AAID	<u>  Length</u>	<u>Length</u>	Score	Probability
23469691_c3_111	946	6168	179	540	7	
Protein name			*	Locus	Name	Acc#
Description						
NO-HIT						
MO-HII						
ODE Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
ORF Name 23515762_f2_52		<u>AAID</u>	<u>Length</u>	Length		
			<u>Length</u>	Length	Score  S Name	Probability  Acc#
235157.62 <u></u> £2 <u></u> 52			<u>Length</u>	Length		
23515762f252 Protein name			<u>Length</u>	Length		
23515762_f2_52 Protein name  Description			Length	Length 279 Locus	s Name	Acc#
23515762_f2_52 Protein name  Description			<u>Length</u>	Length		
23515762_f2_52Protein name  Description  NO-HIT	947 NTID	6169	Length  92  NT  Length	Length  Locus	s Name	Acc#
23515762_f2_52Protein name  Description  NO-HIT  ORF Name	947 NTID	6169	Length  92  NT  Length	Length Locus  AA Length	Score	Acc# Probability
Protein name  Description  NO-HIT  ORF Name  23.984402c182	947  NTID 948	6169	Length  92  NT  Length	Length Locus  AA Length	Score  488  Name	Acc# Probability  8.9e-44
Protein name  Description  NO-HIT  ORF Name  23.984402_c1_82  Protein name	NTID 948 (RagA)	AAID [6170	Length  NT Length  711	Length  Locus  AA  Length  Locu  gp:PGI	Score 488 S Name	Acc#  Probability  8.9e-44  Acc#  AJ130872

	NEETD AAID	NT AA Score	Probability
ORF Name	NTID AAID		
24640675_c2_99	949 6171	511 1536	~
Protein name		Locus Name	Acc#
Description			
NO-HIT			
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
24806300 <u></u> £2 <u></u> 40	950 6172	267 804	
Protein name		Locus Name	Acc#
Description			
NO-HIT			v
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	· · ·
25706687 <u></u> c1 <u></u> 83	951 6173	1248 162	2.3e-17
Protein name		Locus Name	Acc#
receptor antigen	(RagA)	gp:PGI130872	AJ130872
Description			
Porphyromonas gir immunodominant 55k	ngivalis W50 recepto «Da antigen.	r antigen (rag) locus en	codinga major
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length Score	FIODADIIICY
26360636 <u></u> f1 <u></u> 5	952 6174	201 606 208	8.0e-17
Protein name		Locus Name	Acc#
		gp:AHU56832	U56832
Description		_	
	nila FK506 binding p	protein (fkpA) gene, comp	oletecds in 3.9
kb fragment.			· · · · · · · · · · · · · · · · · · ·

ķ,	House House
8	\$200
il batter	i
**)	11000
81	akti
th1	444
F-15	
55,5144	1
47	HR.
340435	
371	Post.
4	
H	
#	77
Hord	in in
. 7	235
100	hills Nac
100	845 347
ä,	32
	ROE
781	Harr
	Marie Marie
- 1	
10	ES.
Ú,	iril sef

ORF Name NTID AAID Length Leng	Score Propagative
2813912_c3_110	508 1.3e-48
Protein name	ocus Name Acc#
sp:	YHAM_ECOLI P42626
Description	·
HYPOTHETICAL 19.4 KD PROTEIN IN EXUR-TDCC INTERGEN	IC REGION (F188)
NT AZ	Score Propaditivy
ORF Name NTID AAID Length Leng	jth —
3417677_c1_85 954 6176 354 1065	171 3.3e-12
Process name	ocus Name Acc#
KIAA0879 protein gp:	AB020686 AB020686
Description	
The second make for KIAA0879 protein, complete c	as.
Homo sapiens mRNA for KIAA0879 protein, complete o	
NT A	A Score Probability
ORF Name NTID AAID Length Len	A Score Probability
ORF Name         NTID         AAID         Length         Len           3.7.9.25tl10         955         6177         63         192	Score Probability  54  0.020
ORF Name  NTID  AAID  Length  Length  1955  Protein name  NT A  AAID  192	Score Probability  54 0.020  Socus Name Acc#
ORF Name         NTID         AAID         Length         Length           3.7.925tl10         955         6177         63         192           Protein name         I         gp	Score Probability  54 0.020  Socus Name Acc#
ORF Name         NTID         AAID         Length         Length           3.7.9.25tl10	Score Probability  54 0.020  ocus Name Acc#  AFSCR X70080
ORF Name  NTID  AAID  Length Length Length  Protein name  Description  A.franciscana Scr gene (homologue of Drosophila Se	Score Probability  54 0.020  Cocus Name Acc#  AFSCR X70080  Ex combs reduced).
ORF Name  NTID  AAID  Length L	Score Probability  54 0.020  cocus Name Acc#  AFSCR X70080  ex combs reduced).
ORF Name  NTID  AAID  Length	Score Probability  54 0.020  Cocus Name Acc#  AFSCR X70080  Ex combs reduced).
ORF Name         NTID         AAID         Length         Length           3.7925fl10         955         6177         63         192           Protein name         I         gp           Description         A.franciscana Scr gene (homologue of Drosophila Scrophila Scro	Score Probability  54 0.020  Cocus Name Acc#  AFSCR X70080  Ex combs reduced).
ORF Name         NTID         AAID         Length         Length           3.7.925tl10         955         6177         63         192           Protein name         Image: Second of the protein name of t	Score Probability  54 0.020  Socus Name Acc#  AFSCR X70080  Ex combs reduced).  A Score Probability
ORF Name  NTID  AAID  Length  Length  Length  Length  Length  Protein name  Description  A.franciscana Scr gene (homologue of Drosophila Scr  ORF Name  NTID  AAID  Length  Length  AAID  Length  Length  AAID  AAID  Length  Length  AAID  Length  Length  AAID  AAID  Length  Length  AAID  AA	Score Probability  54 0.020  Socus Name Acc#  AFSCR X70080  Ex combs reduced).  A Score Probability

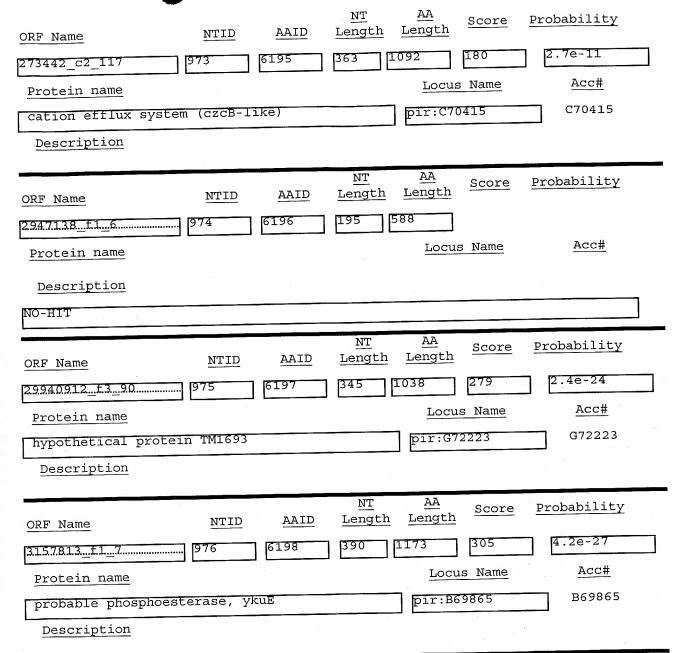
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score Probability
4742812 f3 64	957   6179	378 1137	7.2e-50
			us Name Acc#
Protein name  [hypothetical protein protein protein protein name]	rotein		H132745 AJ132745
	.000111		
Description	aliana hypothetical p	orotein, clone EMG	9a29.
Arabidopsis tha	allana nypochecical p	,2000227,	
ORF Name	NTID AAID	NT AA Length Lengt	<u> Score</u> <u>Probability</u>
4804562_c2_103	958 6180	452 1359	156 2.6e-11
Protein name		Loc	cus Name Acc#
putative outer	membrane porin	gp:Al	030977
Description			
ton mombrane	e glutamyl tRNA synth porin (ompA), unknowr B protein (viuB) gene	n protein, vibriok	e, partial cds;putative pactinreceptor precursor and VibF (vibF) gene,
r		NT AA	. Score Probability
ORF Name	NTID AAII	D <u>Length</u> <u>Lengt</u>	h
	1		180 7.4e-14
ORF Name  48.75.035_c1_80  Protein name	959 6181	Length Lengt  193  582  Loo	180 7.4e-14  cus Name Acc#
ORF Name  48.75.035_c1_80  Protein name	1	Length Lengt  193  582  Loo	180 7.4e-14
ORF Name  48.75.035c180  Protein name  RNA polymerase  Description	sigma factor SigZ-li	Length Length  193   582  Louike protein   gp:A	180 7.4e-14  cus Name Acc#  F137263 AF137263
ORF Name  48.7.5.0.3.5	sigma factor SigZ-li	Length Length  193 582  Louike protein gp:A	180 7.4e-14  cus Name Acc#
ORF Name  48.75.035_c1_80  Protein name  RNA polymerase  Description  Bacteroides the gene cluster, a	sigma factor SigZ-li	Length Length  193 582  Locate protein gp:A  ibosomal protein gma factorSigZ-lil	h  Tous Name  Acc#  F137263  AF137263  AF137263  S16-likeprotein, fucose ke protein (sigZ) genes,  Score Probability
ORF Name  48.75.035_c1_80  Protein name  RNA polymerase  Description  Bacteroides the gene cluster, a	sigma factor SigZ-li	Length Length  193 582  Locative protein gp:A  ibosomal protein gma factorSigZ-li	180 7.4e-14  Cus Name Acc# F137263 AF137263  S16-likeprotein, fucose Re protein (sigZ) genes,  Ch Score Probability
ORF Name  48.75.035c180  Protein name  RNA polymerase  Description  Bacteroides the gene cluster, a complete cds.	sigma factor SigZ-linetaiotaomicron 30S rind RNA polymerase signal NTID AAI	Length Length  193 582  Locate protein gp:A  ibosomal protein gma factorSigZ-lil	h  Tous Name  Acc#  F137263  AF137263  AF137263  S16-likeprotein, fucose ke protein (sigZ) genes,  Score Probability
ORF Name  48.75.035_c1_80  Protein name  RNA polymerase  Description  Bacteroides the gene cluster, a complete cds.  ORF Name	sigma factor SigZ-linetaiotaomicron 30S rind RNA polymerase signal NTID AAI	Length Length  193 582  Locate protein gp:A  ibosomal protein gma factorSigZ-lil  NT AA  D Length Length	h  Tous Name  Acc#  F137263  AF137263  AF137263  S16-likeprotein, fucose ke protein (sigZ) genes,  th  Score Probability
ORF Name  48.75.035c180  Protein name  RNA polymerase  Description  Bacteroides the gene cluster, a complete cds.  ORF Name  48.82.012c195	sigma factor SigZ-linetaiotaomicron 30S rind RNA polymerase signal NTID AAI	Length Length  193 582  Locate protein gp:A  ibosomal protein gma factorSigZ-lil  NT AA  D Length Length  377 1131	h  Tous Name  Acc#  F137263  AF137263
ORF Name  48.75.035c180  Protein name  RNA polymerase  Description  Bacteroides the gene cluster, a complete cds.  ORF Name  48.82.012c195	sigma factor SigZ-linetaiotaomicron 30S rind RNA polymerase signal NTID AAI	Length Length  193 582  Locate protein gp:A  ibosomal protein gma factorSigZ-lil  NT AA  D Length Length  377 1131	180 7.4e-14  Cus Name Acc# F137263 AF137263  S16-likeprotein, fucose ke protein (sigZ) genes,  Ch Score Probability  253 1.4e-19  Cus Name Acc#

		NT AA	Score Probability
ORF Name	NTID AAID	Length Length	
5351507_f2_39	961 6183	378 1137	
Protein name		Locus	Name Acc#
Description			
NO-HIT			
		NT AA	
ORF Name	NTID AAID	Length Length	Score Probability
5.8.8.1.8.7.7c310.7	962 6184	352 1059	1.3e-07
Protein name		Locus	Name Acc#
transmembrane sen	sor	gp:AF05	AF051691
Description			
Pseudomonas aerug	inosa stress factor sor (fiuR), and hydro	A (psfA), ECF sign	na factor(fiuI), derophore receptor
(fiuA) genes, comp	olete cds.	11	*
(fiuA) genes, comp	plete cds.	NT AA	
(fiuA) genes, comp	NTID AAID		Score Probability
(fiuA) genes, comp	NTID AAID	NT AA	
(fiuA) genes, comp	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score Probability 209 1.2e-13
ORF Name  970385c3108  Protein name  serine/threonine	NTID AAID	NT AA Length Length  824 2475  Locus	Score Probability  209 1.2e-13  Name Acc#
ORF Name  970385c3108  Protein name  serine/threonine protein	NTID AAID  963 6185	NT AA Length Length  824 2475  Locus	Score Probability  209 1.2e-13  Name Acc#
ORF Name  970385c3108  Protein name  serine/threonine	NTID AAID  963 6185	NT AA Length Length  824 2475  Locus	Score Probability  209 1.2e-13  Name Acc#
ORF Name  970385c3108  Protein name  serine/threonine protein	NTID AAID  963 6185	NT AA Length Length  824 2475  Locus	Score Probability  209 1.2e-13  Name Acc#
ORF Name  9.703.85c310.8  Protein name  serine/threonine protein  Description	NTID AAID  Protein kinase relat  NTID AAID  AAID	NT AA Length Length  824 2475  Locus  Eed pir:H69	Score         Probability           209         1.2e-13           Name         Acc#           064         H69064
ORF Name  970385	NTID AAID  Protein kinase relat  NTID AAID  AAID	NT AA Length Length    824   2475     Locus     Locus     Length   Pir: H69     Length Length     Length Length     297   894	Score Probability  209 1.2e-13  Name Acc#  064 H69064  Score Probability
ORF Name  970385_c3_108  Protein name  serine/threonine protein  Description  ORF Name  1064063_c2_139  Protein name	NTID AAID  Protein kinase relat  NTID AAID  AAID	NT AA Length Length  824 2475  Locus  Locus  NT AA Length Length  297 894  Locus	Score         Probability           209         1.2e-13           Name         Acc#           064         H69064           Score         Probability           126         0.00030           Name         Acc#
ORF Name  9.7.0.3.85c31.08  Protein name  serine/threonine protein  Description  ORF Name  106.40.63c213.9  Protein name	NTID AAID  963 6185  protein kinase relat  NTID AAID  6186	NT AA Length Length  824 2475  Locus  Locus  NT AA Length Length  297 894  Locus	Score         Probability           209         1.2e-13           Name         Acc#           064         H69064           Score         Probability           126         0.00030           Name         Acc#

SEcret	100
à,	
	Ŋ
	FIII)
110	H)
	1
	an gall
Ł	
184	
10cpu-11	
222	nut:
15,000	
	#22
聯	1177
	Hard I
Fig. P	241 241

7			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
10742332_c1_106	965	6187	85	258			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
11836662 <u></u> £2 <u></u> 46	966	6188	659	1980	1330	1.0e-135	
Protein name				Locus	Name	Acc#	
				sp:YFI	_BACSU	P54719	
Description							
			TATA DOOM	TTTNT 3 TNT	73T.VRC 3	'REGION	
HYPOTHETICAL ABC	TRANS PORTER	ALL-BIND	ING PROI	EIN Z IN	GHVDC 3		
		AAID	NT	AA Length	Score	Probability	
ORF Name	NTID		<u>NT</u> Length	AA			
	NTID	AAID	<u>NT</u> Length	AA Length			
ORF Name  16284692_c2_135  Protein name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	-
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
ORF Name  16284692_c2_135  Protein name  Description	NTID	AAID	NT Length	AA Length 195 Locu	Score	Probability	
ORF Name  16284692_c2_135  Protein name  Description	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
ORF Name  16284692_c2_135  Protein name  Description  NO-HIT	NTID 967	<u>AAID</u> 6189	NT Length 64	AA Length  195  Locus	Score S Name	Probability  Acc#	
ORF Name  16284692_c2_135  Protein name  Description  NO-HIT  ORF Name	NTID 967	AAID 6189 AAID	NT Length 64  NT Length	AA Length  Locus  AA Length  1725	Score Score	Probability  Acc#  Probability	
ORF Name  16284692_c2_135  Protein name  Description  NO-HIT  ORF Name  20507827_f3_79	NTID	<u>AAID</u> 6189  AAID	NT Length 64  NT Length	AA Length  Locus  AA Length  1725	Score  Score  1246  S Name	Probability  Acc#  Probability  8.1e-127	

ORF Name	NTID AAID	NT AA Score Probabi	ility
23562802_c3_144	969 6191	421 1266 575 1.0e-	-55
Protein name		Locus Name A	cc#
		sp:SBCD_RHOCA 06	8033
Description			
EXONUCLEASE SBCD H	OMOLOG		
		NT AA Score Probab	ility
ORF Name	NTID AAID	Length Length Score Probab	TITCA
24651557_c2_138	970 6192	380 1143 78 0.03	8
Protein name		Locus Name A	cc#
fibronecton type I	11	gp:HUMFN3A M1	2549
Description			
Human fibronectin	gene type III homol main, exons 6 and 7.	ogy unit corresponding to	*
thecell-binding dom	talli, exolib o and ,.		
ORF Name	NTID AAID	NT AA Score Probab	ility
	971 6193	996 2991 464 1.5e	-84
Protein name		Locus Name A	cc#
probable exonuclea	ıse,	pir:T03465 T0	3465
Description	ase,	pir:T03465 T0	3465
	ase,		3465
Description	ntid AAID	pir:T03465 TO  NT AA Score Probab Length Length	
Description ORF Name		NT AA Score Probab	
Description  ORF Name  25978516c2119	NTID AAID	NT AA Score Probab Length Length 98	
Description  ORF Name  25.97.8516c2119  Protein name	NTID AAID	NT <u>AA</u> Score Probab Length Length	ility
Description  ORF Name  25978516c2119	NTID AAID	NT <u>AA</u> Score Probab Length Length	ility



ORF Name	NTID AAID	NT AA Score Probability Length Length
34173431_f1_5	977 6199	180 543 183 3.6e-14
Protein name		Locus Name Acc#
SigX		gp:AF115334
Description		
(menG), CmaX (cmaX)	escens PpsA (ppsA) g , CrfX (crfX), CmpX complete cds; and u	gene, partial cds; EstX (estX),MenG K (cmpX), SigX (sigX),OprF (oprF), and unknown gene.
ORF Name	NTID AAID	NT AA Score Probability Length Length
34661301_c1_102		1083 3252 394 6.3e-53
Protein name		Locus Name Acc#
	cance protein (acrB)	homolog pir:D70117 D70117
Description		
		NT AA Grand Brobability
ORF Name	NTID AAID	Length Length Score Probability
3.938215 <u></u> c2 <u></u> 118	979 6201	550 1653 384 7.4e-33
Protein name	Sec. 1	Locus Name Acc#
cation efflux (Ac	rB/AcrD/AcrF family)	pir:F70368 F70368
Description		
		NT AA Score Probability
ORF Name	NTID AAID	Length Length
4394642 <u></u> ±3 <u></u> 71	980 6202	152 459
Protein name		Locus Name Acc#
Description		***
NO-HIT		

į.	54
ž,	7
Sept.	ħ
	5555
13:1	COR.
	U
5.00	
i.	181
12	dia.
12	den.
	E-10
-	
25 - Fr 15	10 mm
March of March 15	
35 25 miles	100 Page 100
The street of the street of	

ORF Name	NTID A	<u>NT</u> AID <u>Leng</u> th	AA Length	Score	Probability	
4805286_c1_99	981 6203	486	1461	533	2.7e-50	
Protein name			Locus	Name	Acc#	
acriflavine resi	stance protein (a	crB) homolo	g pir:D70	)117	D70117	
Description						
		NT	AA	Score	Probability	
ORF Name	NTID A	AID Lengt				-
5195317 <u>_c1</u> _101	982 620	430	1293	110	0.0047	]
Protein name			Locus	s Name	Acc#	
			sp:YD4	_HAEIN	P44165	
Description						
HYPOTHETICAL PRO	TEIN HI1340					
		NT	AA	Score	Probability	
ORF Name	NTID A	<u>NT</u> AID Lengt		Score	Probability	
ORF Name		AID Lengt		Score	Probability	
		AID Lengt	h <u>Length</u>	Score s Name	Probability  Acc#	
6.8.53.43.6. <u></u>		AID Lengt	h <u>Length</u>			
6853436c1115 Protein name		AID Lengt	h <u>Length</u>			
Protein name  Description	983 620	AID Lengt	h <u>Length</u>	s Name	Acc#	
Protein name  Description	983 620	AID Lengt	h Length  [486]  Locu		Acc# Probability	
Protein name  Description  NO-HIT	NTID A	AID Lengt    161	h Length  [486]  Locu	s Name	Acc#	
Protein name  Description  NO-HIT  ORF Name	NTID A	AID Lengt    161	Length Locu  AA h Length Li70	s Name	Acc# Probability	
Protein name  Description  NO-HIT  ORF Name  10235877c2250	NTID A	AID Lengt    161	Length Locu  AA h Length Li70	Score  2007  s Name	Acc# Probability  1.8e-207	
Protein name  Description  NO-HIT  ORF Name  10235877c2250  Protein name  putative epimera  Description	NTID A	AID Lengt  NT AID Lengt  6 389	h Length  AA h Length  Locu  gp:AF1	Score 2007 S Name 25164	Acc#  Probability  1.8e-207  Acc#  AF125164	

7 22 10
Pin
1997
32 IIII
511
ł, "J
1
T.
185 119E
P M
7277
77 76
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

ORF Name 10585817 ±3 105	NTID 1985   16	AAID	NT Length	AA Length	Score	Probability
Protein name	L		<u> </u>		Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
1064005f140	986 6	208	632	L899	291	5.0e-23
Protein name				Locus	Name	Acc#
hypothetical protei	n Rv2731			pir:B70	506	B70506
Description						<u>*</u>
			NT	AA		
ORF Name	NTID	AAID	<u> </u>	<u>Length</u>	Score	Probability
10.66.7943f270	987 6	209	113	342	125	5.0e-08
Protein name				Locus	Name	Acc#
HipA protein.				gp:D907	94	
Description					-	;
E.coli genomic DNA,	Kohara clo	ne #303	(34.3-34.	6 min.).		
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
1096.9017 <u></u> c3293	988	210	343	.032	1742	2.2e-179
Protein name				Locus	Name	Acc#
putative epimerase/	dehydratase			gp:AF12	5164	AF125164
Description						
Bacteroides fragili complete sequence; a			ide B (PS	B2) bio	synthesi	islocus,

	NT AA Score Probability
ORF Name	NTID AAID Length Length
11023432_c1_205	989 6211 414 1245 2081 2.7e-215
Protein name	Locus Name Acc#
putative glycosyl	transferase gp:AF125164 AF125164
Description	
Bacteroides frag complete sequence	lis 638R polysaccharide B (PS B2) biosynthesislocus, and unknown genes.
	NT AA Grana Brobability
ORF Name	NTID AAID Length Length Score Probability
1188951_f2_61	990 6212 60 183
Protein name	Locus Name Acc#
Description	
NO-HIT	
110 1111	
110 1111	NT AA Green Brobability
ORF Name	NTID AAID Length Length
	NTID AAID Length Length Score Probability
ORF Name	NTID AAID Length Length Score Probability
ORF Name  12187817f3117  Protein name	NTID AAID Length Length Score Probability 991 6213 162 489 93 0.031
ORF Name  12187817f3117  Protein name  cell cycle progr  Description	NTID AAID Length Length Score Probability    162   489   93   0.031     Locus Name   Acc#     ession restoration 8 protein   gp:AF011794   AF011794
ORF Name  12187817f3117  Protein name  cell cycle progr  Description	NTID AAID Length Length Score Probability  991 6213 162 489 93 0.031  Locus Name Acc#
ORF Name  12187817f3117  Protein name  cell cycle progr  Description  Homo sapiens cel	NTID AAID Length Length Score Probability  991 6213 162 489 93 0.031  Locus Name Acc# ession restoration 8 protein gp:AF011794 AF011794  1 cycle progression restoration 8 protein (CPR8) mRNA,
ORF Name  12187817f3117  Protein name  cell cycle progr  Description  Homo sapiens cel	NTID AAID Length Length Score Probability    162   489   93   0.031     Locus Name   Acc#     ession restoration 8 protein   gp:AF011794   AF011794
ORF Name  12187817f3117  Protein name  cell cycle progr  Description  Homo sapiens cell complete cds.	NTID AAID Length Length Score Probability    162   489   93   0.031     Locus Name   Acc#     Ession restoration 8 protein   gp:AF011794   AF011794     Cycle progression restoration 8 protein (CPR8) mRNA,     NTID   AAID   Length Length   Score   Probability     NTID   Score   Probability     NTID   AAID   Length Length   Score   Probability     NTID   Score   Probability
ORF Name  12187817f3117  Protein name  cell cycle progr  Description  Homo sapiens cell complete cds.  ORF Name	NTID AAID Length Length Score Probability    162   489   93   0.031     Locus Name   Acc#     Ession restoration 8 protein   gp:AF011794   AF011794     Cycle progression restoration 8 protein (CPR8) mRNA,     NTID   AAID   Length Length   Score   Probability     NTID   Score   Probability     NTID   AAID   Length Length   Score   Probability     NTID   Score   Probability
ORF Name  12187817f3117  Protein name  cell cycle progr  Description  Homo sapiens cell complete cds.  ORF Name	NTID AAID Length Length Score Probability    162   489   93   0.031     Locus Name   Acc#     Ession restoration 8 protein   gp:AF011794   AF011794     Cycle progression restoration 8 protein (CPR8)mRNA,     NTID   AAID   Length Length   Score   Probability     186   18
ORF Name  12187817f3117  Protein name  cell cycle progr  Description  Homo sapiens cel complete cds.  ORF Name  12520688c3280  Protein name	NTID AAID Length Length Score Probability    162   489   93   0.031     Locus Name   Acc#     Ession restoration 8 protein   gp:AF011794   AF011794     Cycle progression restoration 8 protein (CPR8)mRNA,     NTID   AAID   Length Length   Score   Probability     186   18

		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	,
13804187_f1_47	993 6215	98 297 84	0.0018
Protein name		Locus Name	Acc#
hypothetical prot	ein	gp:MTH243656	AJ243656
Description			
Methanobacterium M, N, O, P, Q, & C	thermoautotrophicum ORFS 1,2 & 3.	ehbA, B, C, D, E, F, G,	H, I,J, K, L,
		NT AA Score	Probability
ORF Name	NTID AAID	<u>Length</u> <u>Length</u>	
14250637_£3_147	994 6216	387 1164	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
		NT AA G	
ORF Name	NTID AAID	Length Length Score	Probability
14258450 <u></u> £3 <u></u> 118	995 6217	121 366 100	2.2e-05
Protein name		Locus Name	Acc#
hypothetical prot	tein TM1330	pir:F72267	F72267
Description			
	*		
ORF Name	NTID AAID	NT AA Score Length Length	Probability
14275252 <u>f3</u> 162	996 6218	681 2046 1133	7.6e-115
Protein name	*	Locus Name	Acc#
(p)ppGpp synthet		gp:BSU86377	U86377
	ase	3F	
Description	ase	SF 130	
Racillus subtili	s (p)ppGpp synthetas		

15	100
100	1121 1121
100005	
	1210
##	1100
7	74. 130
Harry Co.	No.
Photo:	100
14,	
9	
32	1122.
	HAP HAP
1	1770
- H	
Total Bridge	1255 1255 1256 1256 1256 1256 1256 1256
33	1000 1000 1000 1000 1000 1000 1000 100
33	1255 1255 1256 1256 1256 1256 1256 1256
H. Straff H Straff	The state of the s
H. Straff H Straff	100 100 100 100 100 100 100 100 100 100

ODE Name	NTID A	NT AID Leng		Score	Probability
ORF Name 14648380 ±1_18	<u>1997</u> 621		888		
Protein name	J L		Locus	s Name	Acc#
Description NO-HIT					
NO-HII					
ORF Name	NTID A	NI AID Leng		Score	Probability
14650882 <u></u> £3114	998 622	0 121	366	96	0.00067
Protein name			Locu	s Name	Acc#
hypothetical prote	in PFB0225c		pir:E7	1620	E71620
Description					
		N	T AA		
ORF Name	NTID	AAID Len		Score	Probability
14660952 <u>f3</u> 150	999 622	21 119	360	220	4.3e-18
Protein name				s Name	Acc#
ybeB protein homo.	log iojap:prote	ein	Locu pir:S7		Acc# S77145
ybeB protein homosslr1886:protein sl	log iojap:prote	ein			
ybeB protein homo.	log iojap:prote	ein			
ybeB protein homosslr1886:protein sl:  Description	log iojap:prote r1886	И	pir:S7	Score	
ybeB protein homosslr1886:protein sl:  Description  ORF Name	log iojap:prote :1886 <u>NTID</u>	N AAID Len	pir:S7	Score	S77145
ybeB protein homosslr1886:protein sl:  Description	log iojap:prote :1886 <u>NTID</u>	N AAID Len	pir:S7  T AA  gth Length	Score	S77145 Probability
ybeB protein homosslr1886:protein sl:  Description  ORF Name	log iojap:prote :1886 <u>NTID</u>	N AAID Len	pir:S7  T AA  gth Length	Score	S77145
ybeB protein homoslr1886:protein sl:  Description  ORF Name	log iojap:prote :1886 <u>NTID</u>	N AAID Len	pir:S7  T AA  gth Length	Score	S77145 Probability

Star Se	
1441	HAN MAI
Hagel	ħ
	440
ţ:	
First	
Sec. 125	
1	
ļ,	44
101	
	es.
	193
#	SH:
Harry.	T.
	11511
#4	pjan;
Manager Sec	72. 227
Strait.	

ORF Name NTID A	NT AA Score Probability AID Length Length
15709675_c1_193	3 367 1104 798 2.4e-79
Protein name	Locus Name Acc#
	sp:YS18_MYCTU P71777
Description	~
HYPOTHETICAL 36.3 KD PROTEIN CY2	77.18
	NIII 7.7
ORF Name NTID A	AAID Length Length Score Probability
	24 401 1206 862 4.0e-86
Protein name	Locus Name Acc#
phosphonopyruvate decarboxylase,	fom2 pir:S60212 S60212
Description	
ORF Name NTID A	NT AA Score Probability AAID Length Length
19533462_c3_2781003   622	17.00.07
the of the state o	Locus Name Acc#
<u>Protein name</u>	sp:YBDG_ECOLI
Description	
HYPOTHETICAL 46.6 KD PROTEIN IN	PHEP-NFNB INTERGENIC REGION
HYPOTHETICAL 40.0 KD PROTEIN IN	
ORF Name NTID F	NT AA Score Probability AAID Length Length
1964058_f2_921004 622	26 279 840 716 1.2e-70
Protein name	Locus Name Acc#
	sp:SOJ_BACSU P37522
Description	

			NT	AA Longth	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>	ПОГС	1.1e-106
20736678_c3_296	1005 62	227	204	515	1056	
Protein name				-	s Name	Acc#
putative undecapr	enyl-phosphate	2		gp:AF12	25164	AF125164
Description						
Bacteroides fragi complete sequence;	lis 638R polys and unknown o	sacchari genes.	ide B (P:	S B2) bio	osynthesi	Islocus,
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length		
22689642_c2_249	1006	228	355	1068	465	4.7e-44
Protein name				Locu	s Name	Acc#
putative glycosyl	transferase			gp:AF1	25164	AF125164
Description						
Bacteroides fragi	lis 638R poly	sacchar:	ide B (P	S B2) bi	osynthes:	islocus,
complete sequence;	and unknown	genes.				
complete sequence;	and unknown	genes.	NT	AA	Score	Probability
complete sequence;	and unknown s	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
complete sequence;	and unknown s	genes.	NT Length	AA Length	Score	Probability  3.4e-12
complete sequence;	and unknown s	AAID	<u>NT</u> Length	AA Length 1527 Locu	Score  195  s Name	Probability  3.4e-12  Acc#
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score  195  s Name	Probability  3.4e-12
ORF Name  23490876_c3_289  Protein name  putative flippase  Description	NTID 1007	AAID 229	NT Length 508	AA Length 1527 Locu gp:AF1	Score 195 s Name 25164	Probability  3.4e-12  Acc#  AF125164
ORF Name  23490876C3289  Protein name  putative flippase	NTID  NTID  ONE Of the control of th	AAID 229	NT Length 508	AA Length 1527 Locu gp:AF1	Score 195 s Name 25164	Probability  3.4e-12  Acc#  AF125164
ORF Name  23490876_c3_289  Protein name  putative flippase  Description  Bacteroides frag- complete sequence	NTID  NTID  ONE OF THE PROPERTY OF THE PROPERT	AAID 229 sacchar genes.	NT Length 508	AA Length  1527  Locu  gp:AF1  S B2) bi	Score 195 s Name 25164	Probability  3.4e-12  Acc#  AF125164
ORF Name  23490876C3289  Protein name  putative flippase  Description  Bacteroides frag:	NTID  OTIO  ITIS 638R poly  and unknown  NTID	AAID sacchar genes.  AAID	NT Length 508	AA Length  Locu  gp:AF1  S B2) bi  AA Length	Score  [195 s Name 25164 osynthes	Probability  3.4e-12  Acc#  AF125164  islocus,
ORF Name  23490876_c3_289  Protein name  putative flippase  Description  Bacteroides frag- complete sequence	NTID  ONTID  ONTID  ONTID  ONTID	AAID 229 sacchar genes.	NT Length 508	AA Length  1527  Locu  gp:AF1  S B2) bi  AA Length  1038	Score  195  s Name 25164  osynthes  Score	Probability  3.4e-12  Acc#  AF125164  islocus,  Probability
ORF Name  23490876_c3_289  Protein name  putative flippase  Description  Bacteroides fraggeomplete sequence	NTID  ONTID  ONTID  ONTID  ONTID	AAID sacchar genes.  AAID	NT Length 508	AA Length  1527  Locu  gp:AF1  S B2) bi  AA Length  1038	Score  [195 s Name 25164 osynthes	Probability  3.4e-12  Acc#  AF125164  islocus,
ORF Name  23490876c3289  Protein name  putative flippase  Description  Bacteroides fraggeomplete sequence  ORF Name  23553136c2238	NTID  ONTID  ONTID  ONTID  ONTID	AAID sacchar genes.  AAID	NT Length 508	AA Length  1527  Locu  gp:AF1  S B2) bi  AA Length  1038	Score  195  s Name 25164  osynthes  Score	Probability  3.4e-12  Acc#  AF125164  islocus,  Probability

ORF Name         NTID         AAID         Length         Score         Probability           [23554555 f3 142         [1009]         [6231]         [254]         [765]
23554555_f3_142
Description
NO-HIT
$rac{ ext{NT}}{ ext{ORF Name}}$ $rac{ ext{NTD}}{ ext{NTID}}$ $rac{ ext{AAID}}{ ext{AAID}}$ $rac{ ext{Length}}{ ext{Length}}$ $rac{ ext{Score}}{ ext{Probability}}$
23.5.9.5.1.3.7£3115
Protein name Locus Name Acc#
Description
NO-HIT
ORF Name NTID AAID Length Length Score Probability
23635952 <u>c1</u> 182
Protein name Locus Name Acc#
dolichol-P-glucose synthetase homolog pir:E69322 E69322
Description
NT AA Score Probability
ORF Name NTID AAID Length Length 1376 1.4e-140
236.7.825224.7
Protein name Locus Name Acc#
phosphoenolpyruvate phosphomutase FOM1 pir:S60206
Description
$rac{ ext{NT}}{ ext{ORF Name}}$ $rac{ ext{NTID}}{ ext{NTID}}$ $rac{ ext{AAID}}{ ext{AAID}}$ $rac{ ext{Length}}{ ext{Length}}$ $rac{ ext{Score}}{ ext{Probability}}$
2382882 <u>c2216</u>
Protein name Locus Name Acc#
hypothetical protein pir:S76344 S76344

ř	
h,	
H,H	ij
	#1
:45	1243
P	
F	1
ķ,	201, 201,
H	
100	874, 11,000
	mg;
105	11911
Į,	
21	300
u;	322
5	
¥#	

	<u>Probability</u>
ORF Name NTID AAID Length Length	5.4e-27
24017687_±3_152	
Protein name Locus Name	Acc#
sp:CDSA_HAEIN	
Description	
SYNTHASE)	
NT AA C	
ORF Name NTID AAID Length Length	Probability
24226587_c2_241	1.1e-26
Protein name Locus Name	Acc#
activator protein gp:AF047527	AF047527
Description	
Pseudomonas fluorescens activator protein (mtlR) gene, complet	ecds.
NIT AA	
ORF Name NTID AAID Length Length	Probability
Score	Probability
ORF Name NTID AAID Length Length	Probability  Acc#
ORF Name         NTID         AAID         Length         Length           24398376_t3_148	
ORF Name         NTID         AAID         Length         Length           24398376t3148	
ORF Name         NTID         AAID         Length         Length           24398376_t3_148	
ORF Name         NTID         AAID         Length         Length           24398376_f3_148         1016         6238         107         324           Protein name         Locus Name           Description           NO-HIT           NT         AA         Score	
ORF Name         NTID         AAID         Length         Length           243398376_f3_148         1016         6238         107         324           Protein name         Locus Name           Description           NO-HIT           ORF Name         NTID         AAID         Length         Length         Score	Acc#
ORF Name         NTID         AAID         Length         Length           24398376_f3_148         1016         6238         107         324           Protein name         Locus Name           Description           NO-HIT           ORF Name         NTID         AAID         Length         Length           24413577_f1_44         1017         6239         288         867	Acc# Probability
ORF Name         NTID         AAID         Length         Length           243398376_f3_148         1016         6238         107         324           Protein name         Locus Name           Description           NO-HIT           ORF Name         NTID         AAID         Length         Length         Score	Acc#
ORF Name         NTID         AAID         Length         Length           24398376_f3_148         1016         6238         107         324           Protein name         Locus Name           Description           NO-HIT           ORF Name         NTID         AAID         Length         Length           24413577_f1_44         1017         6239         288         867	Acc# Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24641925_c1_203	1018 62	40	401 1	.206	136	3.8e-06
Protein name				Locu	s Name	Acc#
galactosyltransfe	rase homolog			pir:G6	9465	G69465
Description						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
24806538 <u></u> c3 <u></u> 292	1019 62	41	342	L029	179	2.0e-11
Protein name				Locu	s Name	Acc#
capsular polysaco	haride biosynt	hsis p	rotein	pir:F7	0441	F70441
Description	*					
			NT	AA		
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability
2537575 <u></u> c2 <u></u> 234	1020 62	242	124	375	87	0.011
Protein name	*			Locu	s Name	Acc#
probable membrane YOL019w:hypothetic	e protein	13		pir:S6	6701	S66701
Description	,					
Description						*
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
26204682 <u></u> ±3 <u></u> .128	1021 62	243	326	981	221	3.3e-18
Protein name				Locu	s Name	Acc#
				sp:Y26	6_ARCFU	029973
Description						

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
26367176 c2 217	1022	6244	251	756			
Protein name			<u></u>	Locus	Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
26379677 <u>_</u> £1 <u>_43</u>	1023	6245	300	903	539	6.7e-52	
Protein name				Locus	Name	Acc#	
				sp:YGI2	PSEPU	P31857	
Description							
HYPOTHETICAL 32.4	KD PROTEIN	IN GIDB-	UNCI INT	ERGENIC F	REGION	8	
			NT	AA		Decoloshilitar	
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
ORF Name 29850282t119		<u>AAID</u>	Length		Score	Probability	
			Length	Length 810			
29850282 <u></u> £1 <u></u> 19			Length	Length 810 Locus	216	1.7e-17	
29850282 <u></u> £1 <u></u> 19			Length	Length 810 Locus	216 Name	1.7e-17 Acc#	
29850282f119 Protein name	1024		Length	Length 810 Locus	216 Name	1.7e-17 Acc#	
Protein name  Description	1024		Length	Length  Locus  sp:Y665	216 S Name S HAEIN	1.7e-17 Acc# P44033	
Protein name  Description	1024		Length	Length 810 Locus	216 Name	1.7e-17 Acc#	
Protein name  Description  HYPOTHETICAL PROT	1024 FEIN H10665 NTID	6246	Length  269  NT  Length	Length Locus Sp:Y669	216 S Name S HAEIN	1.7e-17 Acc# P44033	
Protein name  Description  HYPOTHETICAL PROT  ORF Name	1024 FEIN H10665 NTID	AAID	Length  269  NT  Length	Length Locus Sp:Y669  AA Length	216 S Name S HAEIN	1.7e-17 Acc# P44033	
Protein name  Description  HYPOTHETICAL PROT  ORF Name  32636311_f2_83	1024 FEIN H10665 NTID	AAID	Length  269  NT  Length	Length Locus Sp:Y669  AA Length	216 S Name D HAEIN Score	1.7e-17 Acc# P44033 Probability	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
33376906 c3 290		6248		51	351	5.6e-32
Protein name			<u> </u>	Locu	s Name	Acc#
LicDl				gp:AF1	06539	AF106539
Description						
Streptococcus pneu cds; and unknown ge		1 (licD1	) and Lic	D2 (lic	D2) genes	s,complete
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
33406567_f2_82	1027	6249	925 2	:778	129	6.1e-05
Protein name				Locu	s Name	Acc#
115K outer membran protein	e protein p	recursor	:SusC	pir:JC	6027	JC6027
Description				•		
		<u> </u>	NT	ΔΔ	**	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 33839461_c1_200		<u>AAID</u> 6250	<u>Length</u>	<del></del>	Score	Probability
			<u>Length</u>	Length		
3383.9461 <u></u> c1200	1028	6250	<u>Length</u>	Length	348 s Name	1.2e-31
33839461_c1_200 Protein name	1028	6250	<u>Length</u>	Length .017 Locu	348 s Name	1.2e-31 Acc#
Protein name  putative alcohol d	1028  lehydrogenas	6250 e	Length  338	Length .017 Locu	348 s Name	1.2e-31 Acc#
Protein name  putative alcohol d  Description	1028  lehydrogenas	6250 e	Length  338	Length .017 Locu	348 s Name	1.2e-31 Acc#
Protein name  putative alcohol d  Description  Amycolatopsis orie	In 1028  Hehydrogenas  Entalis cosm  NTID	e nid pCZA3	Length  338  82.  NT Length	Length  Locu  gp:CZA	348 s Name	1.2e-31 Acc# AL078635
Protein name  putative alcohol d  Description  Amycolatopsis orie	In 1028  Hehydrogenas  Entalis cosm  NTID	e and pCZA3	Length  338  82.  NT Length	Length  Locu  Gp:CZA  AA  Length	348 s Name 382 Score	Acc# AL078635  Probability
Protein name  Protein name  putative alcohol of Description  Amycolatopsis orie  ORF Name  35396876c3295	in 1028  Ichydrogenas  Ichydro	e and pCZA3	Length  338  82.  NT Length	Length  Locu  Gp:CZA  AA  Length	348 s Name 382 Score 2046 s Name	1.2e-31  Acc#  AL078635  Probability  1.4e-211
Protein name  putative alcohol d  Description  Amycolatopsis orie  ORF Name  353.968.76_c3_295  Protein name	in 1028  Ichydrogenas  Ichydro	e and pCZA3	Length  338  82.  NT Length	Length  Locu  Gp:CZA  AA  Length  Locu  Locu  Locu  Locu  Locu	348 s Name 382 Score 2046 s Name	1.2e-31  Acc#  AL078635  Probability  1.4e-211  Acc#

ORF Name	NTID AAID	NT AA Score Probability	
35401627 c3 288	7 1030   6252	141   426   431   1.9e-40	<del></del>
Protein name		Locus Name Acc#	
WcgF		gp:AF125164 AF125164	Ł
Description			
Bacteroides fragil complete sequence;		ide B (PS B2) biosynthesislocus,	7
complete sequence;	and unknown genes.		
	ATT 10 10 10 10 10 10 10 10 10 10 10 10 10	NT AA Score Probability	*
ORF Name	NTID AAID	Length Length	
36362675_c1_207	1031 6253	197 594 317 2.2e-28	
Protein name		Locus Name Acc#	
		gp:AB008550 AB008550	)
Description			
Pseudomonas aerugi	nosa phage phi CTX,	complete genome sequence.	٦
		7.7)	
ORF Name	NTID AAID	NT AA Length Length Score Probability	
ORF Name 3914025_c3_287		— Score Probability	
		Length Length Score Probability	
3914025 <u></u> c3 <u></u> 28.7		Length Length Score Probability  166 501 181 5.8e-14	
3914025c328.7		Length Length Score Probability  166 501 181 5.8e-14  Locus Name Acc#	
3914025_c3_287 Protein name  unknown  Description	1032 6254	Length Length Score Probability  166 501 181 5.8e-14  Locus Name Acc#	
3914025_c3_287 Protein name  unknown  Description	. 1032 6254 is 638R polysacchar	Length         Length         Score         Probability           166         501         181         5.8e-14           Locus Name         Acc#           gp:AF125164         AF125164	
Protein name  unknown  Description  Bacteroides fragil	. 1032 6254 is 638R polysacchar	Length Length Score Probability  [166   501   181   5.8e-14  Locus Name Acc#  [gp:AF125164   AF125164  Ide B (PS B2) biosynthesislocus,	
Protein name  unknown  Description  Bacteroides fragil	. 1032 6254 is 638R polysacchar	Length         Length         Score         Probability           166         501         181         5.8e-14           Locus Name         Acc#           gp:AF125164         AF125164	
Protein name  unknown  Description  Bacteroides fragil complete sequence;	is 638R polysacchar and unknown genes.  NTID AAID	Length Length Score Probability  Length Length Score Probability  Locus Name Acc#  AF125164  AF125164  AF125164  MT AA Score Probability	
Protein name  unknown  Description  Bacteroides fragil complete sequence;  ORF Name	is 638R polysacchar and unknown genes.  NTID AAID	Length Length Score Probability  Length Length Score Probability  Locus Name Acc#  AF125164  AF125164  AF125164  NT AA Length Length Score Probability	
3914025_c3_28.7  Protein name  Unknown  Description  Bacteroides fragil complete sequence;  ORF Name  3943.753_c2_245  Protein name	is 638R polysacchar and unknown genes.  NTID AAID	Length Length Score Probability  Length Length Score Probability  Locus Name Acc#  AF125164  AF125164  NT AA Length Length Score Probability  296 891 1278 3.3e-130  Locus Name Acc#	
3914025_c3_28.7  Protein name  Unknown  Description  Bacteroides fragil complete sequence;  ORF Name  3943.753_c2_245  Protein name	is 638R polysacchar and unknown genes.  NTID AAID  1033 6255	Length         Length         Score         Probability           166         501         181         5.8e-14           Locus Name         Acc#           gp:AF125164         AF125164           ide B (PS B2) biosynthesislocus,           NT Length         AA Length         Score         Probability           296         891         1278         3.3e-130           Locus Name         Acc#	
Protein name  unknown  Description  Bacteroides fragil complete sequence;  ORF Name  3943.753c2245	is 638R polysacchar and unknown genes.  NTID AAID  1033 6255  e thymidyltransferas	Length         Length         Score         Probability           166         501         181         5.8e-14           Locus Name         Acc#           gp:AF125164         AF125164           ide B (PS B2) biosynthesislocus,           NT Length         AA Length         Score         Probability           296         891         1278         3.3e-130           Locus Name         Acc#	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
3955062_c3_299	1034	6256	265 7	798	918	4.6e-92
Protein name				Locu	s Name	Acc#
unknown				gp:AF1	25164	AF125164
Description						
Bacteroides fragil complete sequence;			ide B (PS	B2) bio	osynthes:	islocus,
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
3991300_c3_258	1035	6257	295	388	405	1.1e-37
Protein name				Locu	s Name	Acc#
stationary phase s	urvival pr	otein Sur	E	pir:A7	0372	A70372
Description						
			NT	AA	1	_ 1 1 1 2 1 .
ORF Name	NTID	AAID	Length	Length	Score	Probability
4157762 <u></u> c2244	. 1036	6258	182	549	95	0.00012
Protein name				Locu	s Name	Acc#
unknown			υ.	gp:AF0	48749	AF048749
Description						
Bacteroides fragil sequence.	is capsula	r polysac	charide k	oiosynth	esis ope	ron, complete
ORF Name	ŅTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
4175255 <u></u> £3 <u></u> 151	1037	6259	680	2043	1389	4.3e-148
Protein name	,			Locu	s Name	Acc#
FtsH2	9			gp:AB0	23310	AB023310
Description						
Cyanidioschyzon me	rolae gene	for FtsH	12, comple	ete cds.		

OPE Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name		260		23	532	3.7e-51
4304812_c2_246	] [1038]		140			
Protein name					Name	Acc#
WcgG				gp:AF12	25164	AF125164
Description						
Bacteroides fragil complete sequence;			de B (PS	B2) bic	synthesi	Islocus,
			ATM	7.7		-
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4803555_c3_297	1039 62	261	198 5	97	996	2.5e-100
Protein name				Locus	Name	Acc#
putative acetyltra	nsferase			gp:AF12	25164	AF125164
Description						
Bacteroides fragil complete sequence;			de B (PS	B2) bio	synthesi	islocus,
compress sequence,	and ammown	,01100.				*
	× × × × × × × × × × × × × × × × × × ×					
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 48.782.77c1192		<u>AAID</u> 262	Length		Score	Probability
			Length	Length		
4878277 <u>c1</u> 192			Length	Length	95 Name	0.0062
4878277 <u>c1</u> 192			Length	Length  18  Locus	95 Name	0.0062 <u>Acc#</u>
4878277 <u>c1192</u> Protein name	1040 62	262	Length  205  6	Length  18  Locus  gp:YP10	95 Name 02KB	0.0062 <u>Acc#</u>
4878277_cl_192Protein name  Description	1040 62	262	Length  205  6	Length  18  Locus  gp:YP10	95 Name 02KB	0.0062  Acc#  AL031866
4878277_cl_192Protein name  Description	1040 62	262	Length  205  6  gion: fr	Length  Locus  gp:YP10  om 1 to	95 Name 02KB	0.0062  Acc#  AL031866
Protein name  Description  Yersinia pestis 10	1040 62  2 kbases unst	able re	Length  205 6  gion: fr  NT  Length	Length  Locus  gp:YP10  om 1 to	95 Name 02KB	0.0062  Acc#  AL031866
Protein name  Description  Yersinia pestis 10  ORF Name	1040 62  2 kbases unst	able re	Length  205 6  gion: fr  NT  Length	Length  Locus  gp:YP10  om 1 to  AA  Length	95 Name 02KB 119443. Score	Acc# AL031866  Probability
Protein name  Description  Yersinia pestis 10  ORF Name  48.97.128c1201	1040 62  2 kbases unst  NTID  1041 62	able re	Length  205 6  gion: fr  NT  Length	Length  Locus  gp:YP10  om 1 to  AA  Length	95 3 Name 02KB 119443. Score 299 3 Name	0.0062  Acc#  AL031866  Probability  1.8e-26
Protein name  Description  Yersinia pestis 10  ORF Name  4897128_c1_201	1040 62  2 kbases unst  NTID  1041 62	able re	Length  205 6  gion: fr  NT  Length	Length  Locus  gp:YP10  om 1 to  AA  Length  97	95 3 Name 02KB 119443. Score 299 3 Name	D.0062  Acc#  AL031866  Probability  1.8e-26  Acc#
Protein name  Description  Yersinia pestis 10  ORF Name  48.97.128_c1_201  Protein name  N-acetylglucosaming	NTID  1041 62  NTID 1041 62	AAID	Length  205 6  gion: fr  NT Length  298 8	Length  Locus  gp:YP10  om 1 to  AA  Length  97  Locus  gp:AB01	95 Name 02KB 119443. Score 299 Name .7355	D.0062  Acc#  AL031866  Probability  1.8e-26  Acc#  AB017355

		NT	AA	Probability
ORF Name	NTID AAI		Length Score	Probability
4897256_c1_183	1042 6264	498 1	497 1152	7.4e-117
Protein name			Locus Name	Acc#
X-His dipeptidase dipeptidase:aminop		line	pir:JU0300	
D:beta-alanyl-hist				
Description				
		NT	AA Score	Probability
ORF Name	NTID AAI	ID <u>Length</u>	Length	
4962760_c2_233	1043 6265	1927	167	6.8e-20
Protein name			Locus Name	Acc#
hypothetical prot	ein		pir:E72310	E72310
Description				
		NT	AA a	
ORF Name	NTID AA	<del></del>	Length Score	Probability
5136411 <u></u> c1 <u></u> 202	1044 6266	362	178	5.4e-11 <sub>41</sub>
Protein name			Locus Name	Acc#
capsular polysaco	haride biosynthes	sis homolog	pir:F70036	F70036
yveQ				
Description				
		NT	AA Score	Probability
ORF Name	NTID AA		Length -	
5213541 <u></u> c3 <u></u> 263	1045 6267	281	297	3.0e-26
Protein name			Locus Name	Acc#
hypothetical prot	ein APE2014		pir:H72504	H72504
Description				

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
5275281 f1 45	1046	6268	440	1323	375	1.3e-38
Protein name			L	Locus	s Name	Acc#
probable membrane- transglycosylase D		c murein		pir:H7	1301	H71301
Description	3	**************************************				
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
6.03.78.01 <u></u> c3 <u></u> 2.76	1047	6269	379	1140		
Protein name				Locus	s Name	Acc#
Description						
по-ніт						2
			NT	AA		D
ORF Name	NTID	AAID	Length	Length	Score	Probability
6648452 <u></u> £2 <u></u> 87	1048	6270	272	819	396	9.6e-37
Protein name				Locus	s Name	Acc#
				sp:KSG	A_MYCCA	P43038
Description					•	
DIMETHYLTRANSFERA	SE)					
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
6.8.3.2.7.5.7 <u></u> £3 <u></u> .14.9	1049	6271	471	1416	665	3.0e-65
Protein name				Locu	s Name	Acc#
Ykok				gp:AB0	13374	AB013374
Description		-				
Bacillus halodura complete cds.	ns C-125 ma	mX, yjdA,	ykoK an	d yvfK ge	enes, par	rtialand

ORF Name	NTID AAII	<u>NT</u> D <u>Length</u> L	AA Score	Probability
6853387_c1_198	1050 6272	367 110	1050	4.8e-106
Protein name			Locus Name	Acc#
PCZA361.5			gp:AOPCZA361	AJ223998
Description		-		
Amycolatopsis orie	entalis cosmid PC2	ZA361.		
		NT	AA .	
ORF Name	NTID AAII	<del></del>	ength Score	Probability
800812_c2_235	1051 6273	379 114	40	
Protein name			Locus Name	Acc#
Description				
NO-HIT				
		NT	AA	_ 1 1 1 1 1 1 1
ORF Name	NTID AAII		ength Score	Probability
824051 <u></u> c1 <u></u> 206	1052 6274	402 120	1943	1.1e-200
Protein name			Locus Name	Acc#
putative aminotran	sferase	9	gp:AF125164	AF125164
Description				
Bacteroides fragil complete sequence;	. – –		32) biosynthesi	slocus,
	1			
ORF Name	NTID AAII	NT Length L	AA Score	Probability
882702 <u></u> c2 <u></u> 237	1053 6275	279 840	114	1.3e-06
Protein name			Locus Name	Acc#
unknown		C	gp:AF068902	AF068902
Description				
Streptococcus pneu (murD), undecaprenyl transferase(murG), (divIB), orotidine-5 orotatephosphoribos	-PP-MurNAc-pentar cell division pro '-decarboxylase F	peptide-UDPGlo ptein DivIB PyrF (pyrF), a	enac Glenac	ls; andunknown

ORF Name	NTID	AAID Leng		Score	Probability
9944428_±3_97	1054 62	76 100	303		
Protein name			Locus	s Name	Acc#
Description					
NO-HIT					
		NT	AA		
ORF Name	NTID	AAID Leng		Score	Probability
24864688 <u></u> c1 <u></u> 7	1055 62	77 77	234		
Protein name			Locus	Name	Acc#
Description					
NO-HIT	1				
		NT	AA		
ORF Name	NTID	AAID Leng		Score	Probability
29317557 <u></u> c2 <u></u> 9	1056 62	78 519	1560	142	6.3e-06
Protein name			Locus	s Name	Acc#
receptor antigen	(RagA)		gp:PGI1	[30872	AJ130872
Description	(			-	
Porphyromonas gi immunodominant 55	-	ceptor antiq	gen (rag) lo	cus enco	odinga major
Inimario do militario 33	a anorgon.			*	-
		NT	AA	_	- 1 1 1 3 1 I
ORF Name	NTID	AAID Leng		Score	Probability
1038461 <u></u> £3 <u></u> 25	1057 62	79 274	825		
Protein name			Locus	Name	Acc#
Description					
NO-HIT			*		

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
16211377 f3 28	   [1058	6280	158 4	77			
Protein name			L	Locus	s Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
16486287 <u>f1</u> 6	1059	6281	151 4	:56			
Protein name		1		Locus	s Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24647938_f1_5	1060	6282	133 4	.02	110	2.2e-06	
Protein name				Locus	Name	Acc#	, .
muramoyl-pentapepti	de carbox	ypeptidas	е	pir:T34	1747	T34747	
Description							
			NT	AA	Cana	Dwobability	
ORF Name	NTID	$\underline{\mathtt{AAID}}$	Length	Length	Score	Probability	
24666313 <u></u> £3 <u></u> 31	1061	6283	948 2	847	218	6.3e-14	
Protein name				Locus	Name	Acc#	
slow myosin heavy c	hain 2	7		gp:GGU8	35023	U85023	
Description						- × ×	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4100885_£3_26	1062	6284	316	951			
Protein name				Locus	s Name	Acc#	
Description						To:	
тін-ои			3				]
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4119637£18	1063	6285	215	548	115	0.00017	
Protein name				Locus	s Name	Acc#	
hypothetical pr	otein jhp0052			pir:F7	1980	F71980	
Description				7			
			NT	AA			
			414		Caoro	Translandad 1 d 4	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
ORF Name 4900252t11		<u>AAID</u>		Length 795	score	Probability	
				95	S Name	Acc#	
4900252 <u>f1</u> 1				95	*		2
4900252_f1_1				95	*		*
4900252_f1_1 Protein name Description			[264 ] [7	Locus	*	Acc#	, a
4900252_f1_1 Protein name Description				95	*		, a
Protein name  Description  NO-HIT		6286	NT Length	Locus AA	s Name	Acc#	*
Protein name  Description  NO-HIT  ORF Name		6286 AAID	NT Length	Locus  AA  Length	s Name	Acc#	*
490.0252t11  Protein name  Description  NO-HIT  ORF Name  5946.032t216		6286 AAID	NT Length	Locus  AA  Length	Score	Acc# Probability	2 - A

			$\underline{ ext{NT}}$	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length	SCOLE	Probability
10755437_f1_11	1066	6288	168	507	280	1.9e-24
Protein name				Locus	Name	Acc#
				sp:BKDI	R_PSEPU	P42179
Description						<del></del>
BKD OPERON TRANSCR	RIPTIONAL R	EGULATOR				
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
1175211_f3_35	1067	6289	215 6	48	550	4.6e-53
Protein name				Locus	Name	Acc#
inner membrane ABC	transport	er		gp:AF21	3822	AF213822
Description	<del>,</del>					0
Zymomonas mobilis	strain ZM4	fosmid c	Ione 42B3	comple	te segue	ence.
		20020	10110 1220	, сощртс	cc began	
		20020			o beque	
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
	NTID		NT Length	AA		· · · · · · · · · · · · · · · · · · ·
ORF Name	NTID	AAID	NT Length	AA Length		· · · · · · · · · · · · · · · · · · ·
ORF Name 12272127£340	NTID	AAID	NT Length	AA Length	Score	Probability
ORF Name  12272127_f3_40  Protein name	NTID	AAID	NT Length	AA Length	Score	Probability
ORF Name  12272127_f3_40  Protein name  Description	NTID	AAID	NT Length	AA Length	Score	Probability
ORF Name  12272127_f3_40  Protein name  Description	NTID	AAID	NT Length	AA Length	Score	Probability
ORF Name  12272127f340  Protein name  Description  NO-HIT	NTID  NTID	<u>AAID</u> 6290	NT Length  NT Length  Length	AA Length  23 Locus	Score S Name	Probability  Acc#
ORF Name  12272127_f3_40  Protein name  Description  NO-HIT  ORF Name	NTID  NTID	AAID 6290 AAID	NT Length  NT Length  Length	AA Length  Locus  AA Length  AA Length	Score Score	Probability  Acc#  Probability
ORF Name  12272127_f3_40  Protein name  Description  NO-HIT  ORF Name  1423427_f1_4	NTID  NTID	AAID 6290 AAID	NT Length  NT Length  Length	AA Length  Locus  AA Length  AA Length	Score  Score  871  Name	Probability  Acc#  Probability  4.4e-87
ORF Name  12272127_f3_40  Protein name  Description  NO-HIT  ORF Name  1423427_f1_4	NTID  NTID	AAID 6290 AAID	NT Length  NT Length  Length	AA Length  Locus  AA Length  Locus  Locus  Locus  Locus	Score  Score  871  Name	Probability  Acc#  Probability  4.4e-87  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
15657687_f1_13	1070	6292	185	558	373	2.6e-34	
Protein name				Locus	s Name	Acc#	
				sp:YBD	M_ECOLI	P77174	
Description							
HYPOTHETICAL 23.9	KD PROTEIN	IN CSTA-	DSBG INT	ERGENIC F	REGION		1
			NT(T)	7.7			J
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
15900317_f1_5	1071	6293	337	1014	315	1.6e-27	
Protein name				Locus	Name	Acc#	
NrpB	a a			gp:PMU4	16488	U46488	
Description				- L			
Proteus mirabilis l	NrpS (nrpS	) gene, p	artial co	ds. NrpU	(nrpU).	NrpT(nrpT).	1
NrpA (nrpA), NrpB (1	nrpB), Nrp	G (nrpG)		_			*
NrpA (nrpA), NrpB (n	nrpB), Nrp	G (nrpG)		_	enes, cor	mplete cds.	×
NrpA (nrpA), NrpB (n	nrpB), Nrpo	G (nrpG)	and IrpP	(irpP)ge			*
	NTID	_	NT Length	(irpP)ge	enes, cor	mplete cds.	
ORF Name	NTID	AAID	NT Length	(irpP)ge  AA  Length	Score	Probability	
ORF Name 20.98.0080£115	NTID	AAID	NT Length	(irpP)ge  AA  Length	Score  135 Name	Probability 4.0e-08	
ORF Name  20980080f115  Protein name	NTID	AAID	NT Length	AA Length 354 Locus	Score  135 Name	Probability  4.0e-08  Acc#	
ORF Name  20980080_f1_15  Protein name  60kDa protein	NTID	<u>AAID</u> 6294	NT Length	AA Length Locus Locus gp:AB00	Score  [135] S Name  [04560]	Probability  4.0e-08  Acc#  AB004560	
ORF Name  20980080_f1_15  Protein name  60kDa protein  Description	NTID	<u>AAID</u> 6294	NT Length  [117]	AA Length Locus  Locus  gp:AB00	Score  [135] S Name  [04560]	Probability  4.0e-08  Acc#  AB004560	
ORF Name  20980080_f1_15  Protein name  60kDa protein  Description	NTID	<u>AAID</u> 6294	NT Length	AA Length Locus  John Comple	Score  [135] S Name  [04560]	Probability  4.0e-08  Acc#  AB004560	
ORF Name  20980080_f1_15  Protein name  60kDa protein  Description  Porphyromonas ging:	NTID 1072	AAID 6294 for 60kD	NT Length  To protein NT Length  Length	AA Length Locus  John Comple	Score  135 Name 04560	Probability  4.0e-08  Acc#  AB004560	
ORF Name  20.980.080E115  Protein name  60kDa protein  Description  Porphyromonas ging:  ORF Name	NTID 1072  ivalis DNA NTID	AAID  for 60kD	NT Length  To protein NT Length  Length	AA Length  Locus  a, comple  AA Length  1, comple	Score  135 Name 04560	Probability  4.0e-08  Acc#  AB004560	
ORF Name  20.980.080f115  Protein name  60kDa protein  Description  Porphyromonas ging:  ORF Name  226.5.70.52c154	NTID 1072  ivalis DNA NTID	AAID  for 60kD	NT Length  To protein NT Length  Length	AA Length  Locus  a, comple  AA Length  1, comple	Score  135 S Name 04560 Score	Probability  4.0e-08  Acc#  AB004560  Probability	

		$\frac{\text{NT}}{\text{Score}}$	Probability
ORF Name	NTID AAID	<u>Length</u> <u>Length</u>	
22928450_f1_16	1074 6296	60 183	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
		NTT 7.7	
ORF Name	NTID AAID	NT AA Score Length Length	Probability
23.556577 <u></u> £2 <u></u> 25	1075 6297	1305 537	1.2e-59
Protein name		Locus Name	Acc#
		sp:YBDN_ECOLI	P77216
Description			
HYPOTHETICAL 47.8	KD PROTEIN IN CSTA-	-DSBG INTERGENIC REGION	
		7.7	
ORF Name	NTID AAID	NT AA Length Length Score	Probability
23.86.7917c288	1076 6298	96 291 84	0.0043
Protein name		Locus Name	Acc#
MHC class II alpha	a chain	gp:AF091557	AF091557
Description		× <sup>2</sup>	
	enschi MHC class II *01 allele), complet	alpha chain MHC-Auha-DAA ce cds.	1
	6	$\frac{\text{NT}}{\text{NT}}$ , $\frac{\text{AA}}{\text{A}}$ , Score	Probability
ORF Name	NTID AAID	Length Length	,
24431537 <u> </u>	1077 6299	108 327	
Protein name		Locus Name	Acc#
Description			
NO-HIT	-		

ORF Name	NTID AAID	NT AA Length Length Score	Probability
24492078_£1_1	1078 6300	307 924	
<u>Protein name</u>		Locus Name	Acc#
Description			
NO-HIT	***		
		3.77	
ORF Name	NTID AAID	NT AA Length Length Score	Probability
29580387 <u></u> £3 <u></u> 38	. 1079 6301	317 954 142	1.7e-07
Protein name		Locus Name	Acc#
pobR regulator		gp:PSEY18527	Y18527
Description			
Pseudomonas sp. pol	bA, pobR, pcaQ, pca	H and pcaG genes.	
7		2.7	
ORF Name	NTID AAID	NT AA Score	Probability
31691875 <u></u> £2 <u></u> 19	. 1080 6302	214 645 166	4.7e-12
<u>Protein name</u>		Locus Name	Acc#
		gp:LIINLC	Y07639
Description			
L.ivanovii 23S rRNA genes.	A, 5S rRNA, tRNA-As	n, tRNA-Thr, ORF Z, inID	, andinIC
ADD 17		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
33399057 <u></u> c1 <u></u> 52	6303	288 867 159	1.5e-09
Protein name		Locus Name	Acc#
		sp:LCRF_YERPE	P28808
Description			- (1)

			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
35650462_£3_43	1082	6304	200 6	03	272	4.8e-23	
Protein name				Locus	s Name	Acc#	
60kDa protein				gp:AB00	04560	AB00456	0
Description				•			
Porphyromonas gi	ngivalis DNA	for 60kDa	a protein	, comple	te cds.		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4069180_f1_14	1083	6305	197 5	94			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 4.9.5.3.58.6c289		<u>AAID</u>	Length		Score 92	Probability	
			Length	Length			
4953586 <u></u> c2 <u></u> 89	1084		Length	Length	92 Name	0.0026	6
4.953586c289 Protein name	1084		Length	Length 60 Locus	92 Name	0.0026 <u>Acc#</u>	6
4953586_c2_89 Protein name lipase precursor	1084	6306	Length  119  3	Locus  Gp:AF05	92 Name 3006	0.0026 <u>Acc#</u> AF05300	6
4953586_c2_89 Protein name lipase precursor Description	1084	6306	Length  [119] [3]  Sursor (g	Length  Locus  gp:AF05  eh1) gen	92 Name 3006	0.0026 <u>Acc#</u> AF05300	6
4953586_c2_89 Protein name lipase precursor Description	1084	6306	Length  119  3	Locus  Gp:AF05	92 Name 3006	0.0026 <u>Acc#</u> AF05300	6
4.953586c289 Protein name  lipase precursor  Description  Staphylococcus e	pidermidis li	pase prec	Length  119 3  cursor (go  NT  Length	Length  Locus  gp:AF05  eh1) gen	92 Name 3006 e, compl	0.0026 Acc# AF05300	6
4.953586c289  Protein name  lipase precursor  Description  Staphylococcus e	pidermidis li	pase prec	Length  119 3  cursor (go  NT  Length	Length  Locus  gp:AF05  eh1) gen  AA  Length	92 Name 3006 e, complo	0.0026  Acc#  AF05300  etecds.  Probability	6
4.953.586c289	pidermidis li	pase prec	Length  119 3  cursor (go  NT  Length	Length  Locus  Gp:AF05  eh1) gen  AA  Length	Name 3006 e, comple Score 265 Name	0.0026  Acc#  AF05300  etecds.  Probability  7.3e-23	6
4.953.586c289	pidermidis li	pase prec	Length  119 3  cursor (go  NT  Length	Length  Locus  gp:AF05  eh1) gen  AA  Length  Length	Name 3006 e, comple Score 265 Name	0.0026  Acc#  AF05300  etecds.  Probability  7.3e-23  Acc#	6

ORF Name	NTID AAID	NT AA Length Length Score Probability
5260317_c2_80	1086 6308	61 [186 [54 [0.042]
Protein name		Locus Name Acc#
pqqG protein		pir:B55527 B55527
Description		
ORF Name	NTID AAID	NT AA Score Probability Length Length
7.0.6.5.8.0.2±328	. 1087 6309	238 717 81 0.0088
Protein name		Locus Name Acc#
hypothetical prote	in MTH1102	pir:F69013 F69013
Description		
ORF Name	NTID AAID	NT AA Score Probability Length Length
9963202 <u></u> 23 <u></u> 101	. 1088 6310	511 1536 373 1.2e-32
Protein name		Locus Name Acc#
sensory transducti sll0474:protein sll	on histidine kinase 0474:protein sll047	
Description		* * *
,		
ORF Name	NTID AAID	NT AA Score Probability
11723417 <u>_</u> c2_40	. 1089 6311	484 1455 121 2.8e-06
Protein name		Locus Name Acc#
unknown		gp:U96771 U96771
Description		
Prevotella bryanti mannanase genes, co		cturonase,B-1,4-endoglucanase, and nowngenes.

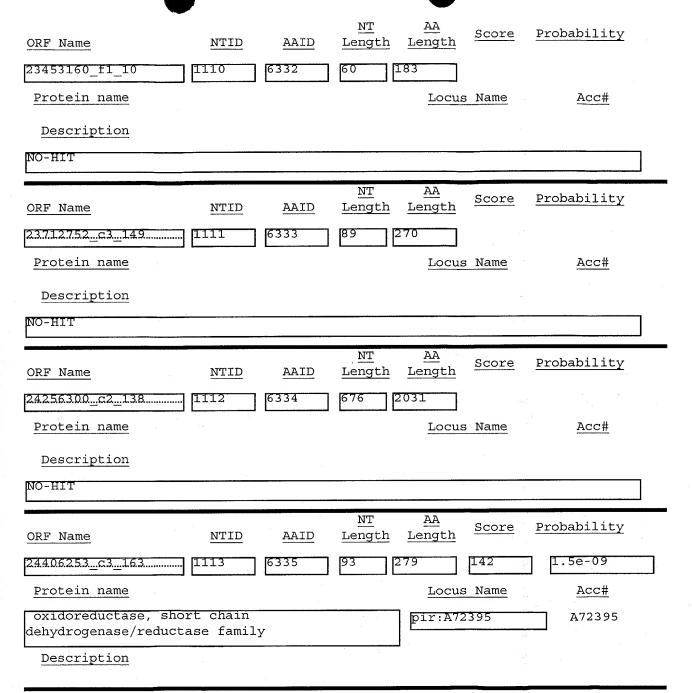
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
16601526_c1_38	1090	6312	1089	3270	801	7.4e-91
Protein name				Locus	Name	Acc#
receptor antigen	(RagA)			gp:PGI1	.30872	AJ130872
Description			-	1)1	······································	
Porphyromonas gi immunodominant 55	_	receptor	antigen	(rag) lo	cus enc	odinga major
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
20507937_c2_42	1091	6313	542	1629	138	3.6e-07
Protein name				Locus	Name	Acc#
unknown		*		gp:U967	71	U96771
Description						· · · · · · · · · · · · · · · · · · ·
Prevotella bryan mannanase genes,	_				naograce	andse, and
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
	NTID		<u>NT</u> Length	AA	Score	Probability 0.024
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length		
ORF Name 21678137t327	NTID	AAID	<u>NT</u> Length	AA Length	101 Name	0.024
ORF Name  21678137_t3_27  Protein name	NTID	AAID	<u>NT</u> Length	AA Length 1263 Locus	101 Name	0.024 <u>Acc#</u>
ORF Name  21678137_t3_27  Protein name  hypothetical prof	NTID	AAID	<u>NT</u> Length	AA Length 1263 Locus	101 Name	0.024 <u>Acc#</u> B69988
ORF Name  216.7813.7_t3_27  Protein name  hypothetical prof	NTID 1092 tein ytap	<u>AAID</u> 6314	NT Length 420	Length  Locus  pir:B69  AA  Length	101 Name 988	0.024 <u>Acc#</u> B69988
ORF Name  216.78137_t3_27  Protein name  hypothetical prot  Description  ORF Name	NTID 1092 tein ytap	AAID 6314 AAID	NT Length 420	Length  Locus  pir:B69  AA  Length	Name 988 Score	0.024 Acc# B69988 Probability
ORF Name  216.7813.7_t3_27  Protein name  hypothetical prof  Description  ORF Name  244.91512_c3_49	NTID 1092 tein ytap	AAID  AAID	NT Length 420	Length  Locus  pir:B69  AA  Length	Name 988 Score 147 Name	0.024 Acc# B69988 Probability 5.0e-07
ORF Name  216.78137_t3_27  Protein name  hypothetical prof  Description  ORF Name  244.91512_c3_49  Protein name	NTID 1092 tein ytap	AAID  AAID	NT Length 420	Length  Locus  Pir:B69  AA  Length  Locus  Locus	Name 988 Score 147 Name	0.024  Acc#  B69988  Probability  5.0e-07  Acc#

ORF Name 35351583_f3_34  Protein name	NTID AAID Length Length  1094  6316  71  Locus Name  ACC#
Description	
NO-HIT	7.7
ORF Name 4120307f326  Protein name hypothetical pro	Locus Name Acci
Description	
ORF Name  4144515f217  Protein name  Description  NO-HIT	NT AA Score Probability Length Length  1096 6318 70 213  Locus Name Acc#
ORF Name 781932_c1_35  Protein name	NT AA Score Probability  NTID AAID Length Length 543 2.7e-85  Locus Name Acc#  brane protein precursor:SusC pir:JC6027 JC6027

ORF Name	NTID AAID	NT AA Score Length Length 614	Probability  [9.2e-68
22697711_c3_9	1098 6320	329 990 614 Locus Name	Acc#
Protein name		gp:BNRNANASE	D28493
neuraminidase pred	cursor	gp . biver. 1	-
Description  Bacteroides fragi	lis nanH gene for ne	uraminidase, complete	cds .
		NT AA Scor	
ORF Name	NTID AAID	Length Length	
11182950_c2_139	1099 6321	248 747 619	2.2e-60
Protein name		Locus Name	Acc#
TruB		gp:AF169967	AF169967
			. 00
	ohnsoniae LeuS (leuS	gene, partial cds; and Jene (trub) gene	nd Fjo12(fjo12),
I Fiavobacceriam J			
FtsX (ftsX), Fjol	3 (fjo13), BacA (bac	A), and Trub (trub)gen	es, compress tar
FtsX (ftsX), Fjol.	3 (I]013), Back (200	NT AA Sco Length Length	
FtsX (ftsX), Fjol.	NTID AAID	NT AA SCO	
ORF Name  11956503_c1_120	NTID AAID	NT AA Sco Length Length	re Probability
FtsX (ftsX), Fjol.	NTID AAID	NT AA Sco Length Length  142 429 110 Locus Nam	re Probability  1.9e-06  Acc#
ORF Name  11956503_c1_120  Protein name	NTID AAID	NT AA Sco Length Length  142 429 110	re Probability  1.9e-06  Acc#
ORF Name  11.956.503c1120  Protein name  Description	NTID AAID  1100 6322	NT AA Sco Length Length 110 Locus Nam Sp:RNPA_BOR	re Probability  1.9e-06  Acc#
ORF Name  11.956.503c1120  Protein name  Description	NTID AAID  1100 6322	NT AA Sco Length Length 110 Locus Nam Sp:RNPA_BOR	re Probability  1.9e-06  ne Acc# P50069
ORF Name  11.956.503c1120  Protein name  Description	NTID AAID  NTID 6322  PROTEIN COMPONENT, (I	NT AA SCO Length Length  142 429 110  Locus Nam  Sp:RNPA_BOR  PROTEIN C5) (RNASE P)  NT AA SCO	re Probability  1.9e-06  Acc#
ORF Name  11.956.503_cl120  Protein name  Description  RIBONUCLEASE P I	NTID AAID  PROTEIN COMPONENT, (F	NT AA Length Length Sco  Locus Nam Sp:RNPA_BOF  PROTEIN C5) (RNASE P)  NT AA Length Length Sco	re Probability  1.9e-06  ne Acc#  RBU P50069
ORF Name  11.956.503cl120  Protein name  Description  RIBONUCLEASE P I	NTID AAID  PROTEIN COMPONENT, (F	NT AA Sco Length Length  142 429 110  Locus Nam  Sp:RNPA_BOF  ROTEIN C5) (RNASE P)  NT AA Length Length Sco	re Probability  1.9e-06  ne Acc# P50069  Dre Probability
ORF Name  11.956.503_cl120  Protein name  Description  RIBONUCLEASE P I	NTID AAID  PROTEIN COMPONENT, (F	NT AA Length Length Sco  Locus Nam Sp:RNPA_BOF  PROTEIN C5) (RNASE P)  NT AA Length Length Sco	re Probability  1.9e-06  ne Acc# P50069  Dre Probability
ORF Name  11.956.503_cl120  Protein name  Description  RIBONUCLEASE P F  ORF Name  13.864.002_fl26	NTID AAID  PROTEIN COMPONENT, (F	NT AA Sco Length Length  142 429 110  Locus Nam  Sp:RNPA_BOF  ROTEIN C5) (RNASE P)  NT AA Length Length Sco	re Probability  1.9e-06  ne Acc# P50069  Dre Probability

			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length		
14882902_c1_108	1102	6324	381 1	146	117	0.00018
Protein name				Locu	s Name	Acc#
sensory transduct	_			pir:S7	7341	S77341
protein slr1837:pr slr1837	otein siri8	37:protei	1			
Description						
			NT	AA		
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability
15109377_c1_119	1103	6325	254 7	765	108	0.0011
Protein name				Locus	s Name	Acc#
				sp:HEM	4_SCHPO	P87214
Description						()
(UROIIIS)		9		1)		A.P.
				<del></del>		
			NТ	ΔΔ		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 20176878c3156		<u>AAID</u>	Length		Score 106	Probability 0.00063
			Length	Length	. ~	
			Length	Length	106 S Name	0.00063
20176878c3156 Protein name			Length	Length  Locus	106 S Name	0.00063 <u>Acc#</u>
20176878_c3_156  Protein name  ATPase subunit 6	1104	6326	Length	Length Locus	106 s Name 40265	0.00063 <u>Acc#</u>
20176878_c3_156  Protein name  ATPase subunit 6  Description	1104	6326	Length  280 8	Length Locus Gp:TCU	106 s Name 40265	0.00063 <u>Acc#</u>
20176878_c3_156  Protein name  ATPase subunit 6  Description	1104	6326	Length	Length Locus	106 s Name 40265	0.00063 <u>Acc#</u>
20176878_c3_156  Protein name  ATPase subunit 6  Description  Trypanosoma cruzi	ATPase sub	6326 unit 6 mRI	Length  280 8  NA, compl	Length Locus Gp:TCUs ete cds	106 S Name 40265	0.00063 <u>Acc#</u> U40265
20176878_c3_156  Protein name  ATPase subunit 6  Description  Trypanosoma cruzi  ORF Name	ATPase sub	unit 6 mRI	Length  280 8  NA, compl	Length Locus  Gp:TCUs  ete cds  AA  Length	106 S Name 40265 Score	0.00063  Acc#  U40265  Probability
Protein name  ATPase subunit 6  Description  Trypanosoma cruzi  ORF Name  204775_c3_153	ATPase sub	unit 6 mRI	Length  280 8  NA, compl	Length Locus  Gp:TCUs  ete cds  AA  Length	106 S Name 40265 Score 466 S Name	0.00063  Acc#  U40265  Probability  3.7e-44
Protein name  ATPase subunit 6  Description  Trypanosoma cruzi  ORF Name  2047.75_c3_153  Protein name	ATPase sub	unit 6 mRI	Length  280 8  NA, compl	Length  Locus  AA  Length  Locus  Locus  Locus	106 S Name 40265 Score 466 S Name	0.00063  Acc#  U40265  Probability  3.7e-44  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
2116037_c1_107  Protein name	1106	6328	76	Locus	s Name	Acc#	
Description						-	
NO-HIT					· · · · · · · · · · · · · · · · · · ·		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
21522150 <u></u> c2 <u></u> 141	1107	6329	437	.314	597	1.4e-110	
Protein name				Locus	Name	Acc#	
				sp:METH	_HAEIN	P43762	
Description						, ,	
ADENOSYLTRANSFERASE	) (ADOMET	SYNTHETA	SE)			м Ф	-
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
ORF Name 23442175c1122	NTID	<u>AAID</u>	Length		<u>Score</u>	Probability 2.8e-108	
			Length	Length			-
23442175 <u></u> c1 <u></u> 122			Length	Length	1071 Name	2.8e-108	
23442175 <u></u> c1 <u></u> 122			Length	Length  Jocus	1071 Name	2.8e-108 <u>Acc#</u>	
23442175_cl_122 Protein name	1108	6330	Length 439	Length  320  Locus  sp:SYY	1071 Name BACST	2.8e-108 <u>Acc#</u>	
23442175_cl_122  Protein name  Description  TYROSYL-TRNA SYNTHE	TASE, (TYR	e330	Length  439  RNA LIGAS  NT	Length  Locus  Sp:SYY  E) (TYRR	1071 Name BACST	2.8e-108  Acc#  P00952	
23442175_cl_122  Protein name  Description  TYROSYL-TRNA SYNTHE	TASE, (TYR	eosineTi	Length  439  RNA LIGAS  NT  Length	Length  Locus  Sp:SYY  E) (TYRR  AA  Length	1071 Name BACST	2.8e-108 <u>Acc#</u>	
Protein name  Description  TYROSYL-TRNA SYNTHE  ORF Name  23447031_c1_109	TASE, (TYR	e330	Length  439  RNA LIGAS  NT  Length	Length  Locus  Sp:SYY  E) (TYRR	1071 Name BACST	2.8e-108  Acc#  P00952	
23442175_cl_122  Protein name  Description  TYROSYL-TRNA SYNTHE	TASE, (TYR	eosineTi	Length  439  RNA LIGAS  NT  Length	Length  Locus  Sp:SYY  E) (TYRR  AA  Length	1071 Name BACST	2.8e-108  Acc#  P00952	
Protein name  Description  TYROSYL-TRNA SYNTHE  ORF Name  23447031_c1_109	TASE, (TYR	eosineTi	Length  439  RNA LIGAS  NT  Length	Length  Locus  Sp:SYY  E) (TYRR  AA  Length	Name BACST Score	2.8e-108  Acc# P00952  Probability	



ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
26605287_c1_114  Protein name	1114	6336	268	Jocus	400 s Name	3.6e-37 Acc#
Trocom name					A_ECOLI	<u> 1100#</u>
Description						
(EC 2.7.1.66)						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
29303427_f1_14	1115	6337	81 2	246	70	0.033
Protein name				Locus	s Name	Acc#
hypothetical prote	ein A635R			pir:T18	3137	T18137
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 29932918c3151		<u>AAID</u>	Length		Score	Probability
***************************************			Length	Length	Score S Name	Probability  Acc#
			Length	Length		
29932918c3151 Protein name			Length	Length		
Protein name  Description			Length	Length	s Name	Acc#
Protein name  Description			Length	Length		Acc# Probability
Protein name  Description	[1116   NTID	6338	Length  85  NT Length	Length  258  Locus	s Name	Acc#
Protein name  Description  NO-HIT  ORF Name	[1116   NTID	AAID	Length  85  NT Length	Length Locus  AA Length	S Name	Acc# Probability
Protein name  Description  NO-HIT  ORF Name  32228430_c3_161  Protein name  choline sulfatase	[1116   NTID	AAID	Length  85  NT Length	Length Locus  AA Length	Score  447  Name	Acc# Probability 3.8e-42
29932918_c3_151  Protein name  Description  NO-HIT  ORF Name  32228430_c3_161  Protein name	NTID	AAID 6339	Length  85  NT Length  527  1	Length Locus  AA Length Locus  Locus	Score  447  Name	Acc#  Probability  3.8e-42  Acc#

ORF Name	NTID AF	NT AID Length	Length Score	Probability
34251637_£2_70	1118   6340		273	
Protein name			Locus Name	Acc#
Description				
NO-HIT				
×		NT	AA Score	Probability
ORF Name	NTID A	Length	Length Door	
35791416 <u></u> c1 <u></u> 105	1119 6341	692	2079 449	2.4e-39
Protein name			Locus Name	Acc#
putative secreted	beta-galactosic	lase	gp:SCF81	AL133171
Description		-		
Streptomyces coel	icolor cosmid F8	31.		
		NT	AA ~	
ORF Name	NTID AF	AID Length	Length Score	Probability
36366552 <u></u> c1 <u></u> 113	1120 6342	79	240 175	2.5e-13
Protein name			Locus Name	Acc#
Fjo13			gp:AF169967	AF169967
			SF	
Description		,		
Description Flavobacterium jo FtsX (ftsX), Fjo13	hnsoniae LeuS (l (fjo13), BacA (	.euS) gene, p (bacA), and T	artial cds; and	Fjo12(fjo12), , complete cds.
Flavobacterium jo	hnsoniae LeuS (] (fjo13), BacA (	LeuS) gene, p (bacA), and T <u>NT</u>	artial cds; and ruB (truB)genes	, complete cds.
Flavobacterium jo	(fjol3), BacA	(bacA), and T	artial cds; and ruB (truB)genes	, complete cds.
Flavobacterium jo FtsX (ftsX), Fjo13	(fjo13), BacA	(bacA), and T  NT  Length	artial cds; and ruB (truB)genes	, complete cds.
Flavobacterium jo FtsX (ftsX), Fjo13 ORF Name	(fjo13), BacA	(bacA), and T  NT  Length	artial cds; and ruB (truB)genes  AA Score	, complete cds.
Flavobacterium jo FtsX (ftsX), Fjo13 ORF Name	(fjo13), BacA	(bacA), and T  NT  Length	artial cds; and ruB (truB)genes  AA Score Length	Probability
Flavobacterium jo FtsX (ftsX), Fjo13  ORF Name  3643.78.93_c3_148  Protein name	(fjo13), BacA	(bacA), and T  NT  Length	artial cds; and ruB (truB)genes  AA Score Length	Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
3937750_c1_110	1122 6	344		.065		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
3942082c2140	1123 6	345	357 1	074	753	1.4e-74
Protein name				Locus	Name	Acc#
S-adenosylmethionin	e tRNA ribo	syltrans	ferase	pir:A72	2360	A72360
Description						
			NT/II	7.7		
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
4102177c3162	1124 6	346	283	352	735	1.1e-72
Protein name				Locus	s Name	Acc#
				sp:KDU	_ERWCH	Q05529
Description						
(5-KETO-4-DEOXYURON	ATE ISOMERA	SE) (DKI	ISOMER <i>A</i>	SE)		
			NTITE	7.7	-	*
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
4119005 <u></u> a3158	1125 6	347	212	39	159	1.6e-11
Protein name				Locus	s Name	Acc#
HI0454				gp:AF1	74390	AF174390
Description						
Haemophilus influen	zae strain	Rd KW20	HI0454 C	gene, par	tial cds	3.

ORF Name 4188438_c3_157 Protein name conserved hypothet Description	1126 634	NT AAID Length	AA Length Score  231 226  Locus Name  pir:G72251	Probability  9.9e-19  Acc#  G72251
ORF Name  4398382_c1_118  Protein name  conserved hypothet  Description	. 1127 634		AA Score Length Score  Locus Name  pir:D70033	Probability  2.1e-34  Acc#  D70033
ORF Name  4757c3152  Protein name  Description  NO-HIT		AAID Length	— Score	Probability  Acc#
ORF Name  486.7142f374  Protein name  hypothetical protein description	. 1129 635	NT AAID Length	AA Score Length 302 Locus Name pir:S73091	Probability  1.1e-38  Acc#  S73091
ORF Name  487.6438_t3_102  Protein name  response regulator  Description  Streptococcus pneur	[1130] [635		1773 162  Locus Name  gp:SPAJ6398	Probability  1.7e-09  Acc#  AJ006398  ystem09.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
Protein name	1131	6353	376	Locus	Name	Acc#
Description NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4954376 <u></u> c2137	1132	6354	355 1	.068		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
5273377 <u></u> c2 <u></u> 132	1133	6355	359 1	.080		
Protein name				Locus	Name	Acc#
Description						
NO-HIT		=				
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
553161c2133	1134	6356	81 2	46		
Protein name				Locus	Name	Acc#
Description						
NO-HIT				*		
			3700	7.7	-	
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
553161 <u></u> c2 <u></u> 135	1135	6357	83 2	52		
Protein name	_			Locus	Name	Acc#
Description						
NO-HIT						
L						

E	
4	1225
Į,	
	20%
\$tt	1000
West.	442
Beerr	No.
£ 200	How the
a.	1117 1117
Щ	
	100 100 100 100 100 100 100 100 100 100
\$57	non.
See See	
;;;	1130
Hard.	

		777	NT	<u>AA</u> Length	Score	Probability	
ORF Name	NTID	AAID	Length		H 2.2	4.7e-06	
5892138_c1_106	1136	6358	409 1	.230	133		
Protein name				Locus	Name	Acc#	
hypothetical prote	in PH0283			pir:D71	.453	D71453	
Description							
×			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
6829630 <u></u> c2 <u></u> 136	. 1137	6359	376	131			
Protein name	*			Locus	s Name	Acc#	
Description							
NO-HIT							
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
6835381 <u>f3</u> 93	1138	6360	65	198		1	
Protein name				Locus	s Name	Acc#	
Description							
NO-HIT	_						
			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
6837828 <u></u> c1117	1139	6361	154	465	326	2.5e-29	
Protein name				Locu	s Name	Acc#	
				sp:HPP	K_PORGI	083019	
Description							
(HPPK) (6-HYDROXY)	METHYL-7,8-	-DIHYDROP'	TERIN PYR	OPHOSPHO	KINASE)	(PPPK)	-

	NULL NATO	<u>NT</u> <u>AA</u> Length Length	Score	Probability
ORF Name	NTID AAID		3.40	1 12 70 07
7315641_c1_111	1140 6362	311 936	140	2.7e-07
Protein name		Locus	s Name	Acc#
ubiquinone/menaquino		pir:F7	2262	F72262
methyltransferase-rel	ated protein			
Description				
		NT AA		
ORF Name	NTID AAID	Length Length	Score	Probability
10978590 <u>t3</u> 64	1141 6363	77 234		
Protein name		Locu	s Name	Acc#
Description			*	. *
NO-HIT	3			
		NT AA	Caoro	Probability
ORF Name	NTID AAID	Length Length	Score	FIODADITICY
14257762f366	1142 6364	110 333		
Protein name		Locu	s Name	Acc#
Description				
NO-HIT				
		NT AA		
ORF Name	NTID AAID	Length Length	Score	Probability
16453180 <u></u> £3 <u></u> 65	1143 6365	150 453	50	0.020
Protein name		Locu	ıs Name	Acc#
WW domain binding p	rotein 5	gp:MMC	J92454	U92454
Description				
Mus musculus WW dom	ain binding prote	in 5 mRNA, partia	l cds.	

ORF_Name  2084768_f3_57  Protein name  Description  NO-HIT	NTID AAID 1144 6366	NT AA Length Length 149 450 Locu	Score S Name	Probability  Acc#	
		NIII NA			
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score	Probability	
23.705002 <u></u> £1 <u></u> 10	1145 6367	746 2241	243	1.2e-29	ĺ
Protein name		Locu	s Name	Acc#	
conserved hypotheti	cal protein ylbK	pir:H6	9874	H69874	
Description					
			· · · · · · · · · · · · · · · · · · ·	*	
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score	Probability	
24422676_t1_15	1146   6368	1023  3072			
Protein name		Locu	s Name	Acc#	
Description					
NO-HIT					
ORF Name	NTID AAID	NT AA Length Length	Score	Probability	
24492125 <u>c1</u> 105	1147 6369	1025 3075	2141	1.2e-221	
Protein name		Locu	s Name	Acc#	
hypothetical protei	n mexF	pir:T3	0830	T30830	
Description	* -		-		
					_
ORF Name	NTID AAID	NT AA Length Length	Score	Probability	
24643800 <u></u> £2 <u></u> 37	1148 6370	258 777	572	2.1e-55	
Protein name	-	Locu	s Name	Acc#	
		sp:YAF	V_ECOLI	Q47679	
Description					
HYPOTHETICAL 28.9 K	D PROTEIN IN DNAQ	-GMHA INTERGENIC	REGION		
				<del>-</del>	

ORF Name		AAID Lei	NT AA ngth Length	Score	Probability
25975307_f1_27  Protein name	[1149] [63]	71 112		s Name	Acc#
Description NO-HIT					
ORF Name 26680340tl16	NTID 2	AAID Lei	NT AA ngth Length	Score	Probability
Protein name Description			Locus	s Name	Acc#
NO-HIT					
ORF Name	NTID 2	-	NT AA ngth Length	Score	Probability
Protein name  Description	1151 63	409	لــــالـــ	485  Name E_ECOLI	3.5e-46 Acc# P24180
ACRIFLAVIN RESISTAN	CE PROTEIN E	PRECURSOR	(ENVC PROTEI	.N)	
ORF Name	NTID Z	_	NT <u>AA</u> ngth Length	Score	Probability
30583162_c2_108 Protein name	1152 637	981		1508 Name	2.3e-166 Acc#
transcription-repair	coupling fa	ctor	gp:AF02	3181	AF023181
Description		1		-	· .
Listeria monocytogen temperature requireme complete cds.					

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
31375817_f2_44	1153	6375	113 3	42	69	0.042	]
Protein name				Locus	s Name	Acc#	
conserved hypothet	tical protei	n AF0188		pir:D69	9273	D69273	
Description							
			NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length	Score	PIODADITICY	
32147151f112	1154	6376	392 1	.179	444	7.8e-42	]
Protein name				Locus	s Name	Acc#	
				sp:NAG	A_VIBCH	032445	
Description				- 8			
DEACETYLASE)						-	
DEACETYLASE)							
DEACETILASE)			NT	AA	Caoro	Drobability	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
		<u>AAID</u>	Length		Score	Probability	
ORF Name			Length	Length			]
ORF Name 3.22826.3.7t236	1155		Length	Length	310 s Name	1.2e-27	
ORF Name 32282637f236  Protein name	1155		Length	Length 591 Locu	310 s Name	1.2e-27 <u>Acc#</u>	]
ORF Name  32282637t236  Protein name  hypothetical prot	1155		Length	Length  Locu  pir:G7	310 s Name 5263	1.2e-27 Acc# G75263	]
ORF Name  32282637t236  Protein name  hypothetical prot	1155		Length	Length 591 Locu	310 s Name	1.2e-27 <u>Acc#</u>	]
ORF Name  32282637t236  Protein name  hypothetical prot  Description	ein NTID	6377	Length  196  NT Length	Length Locu pir:G7	310 s Name 5263	1.2e-27 Acc# G75263	_
ORF Name  32282637t236  Protein name  hypothetical prot  Description  ORF Name	ein NTID	6377 AAID	Length  196  NT Length	Length Locu pir:G7  AA Length	310 s Name 5263 Score	1.2e-27 <u>Acc#</u> G75263  Probability	_
ORF Name  32282637f236  Protein name  hypothetical prot  Description  ORF Name  34181513f126	ein  NTID  [1156	<u>AAID</u>	Length  196  NT Length	Length Locu pir:G7  AA Length	310 s Name 5263 Score 293 s Name	1.2e-27  Acc#  G75263  Probability  4.1e-42	

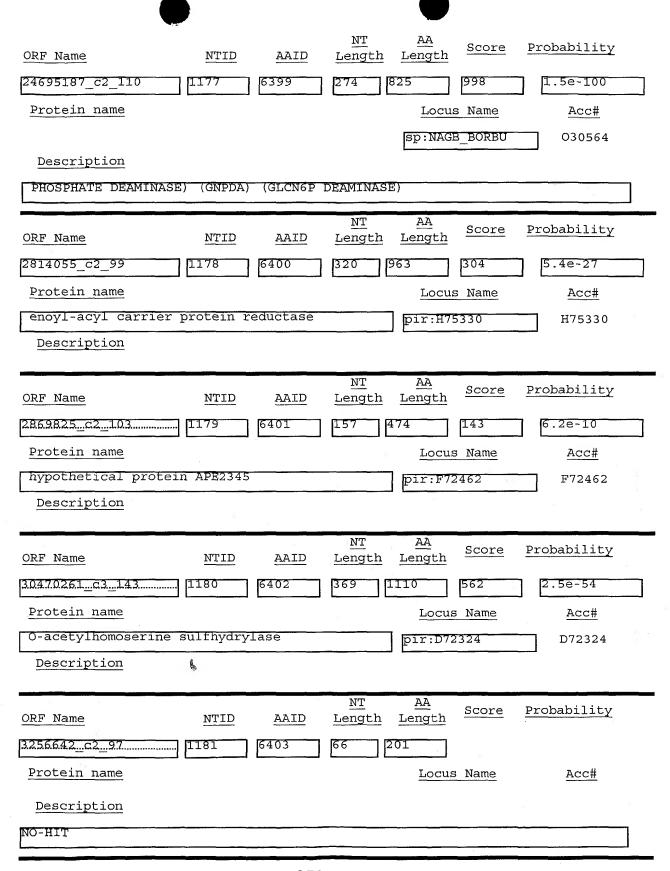
ORF Name	NTID Z	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
34430317_f2_38	1157 63'	79	262 7	89	304	5.4e-27
Protein name				Locus	s Name	Acc#
protein-tyrosine p	phosphatase			gp:AB02	28630	AB028630
Description				,		
Clostridium perfr protein, bacterial nucleotide 2'-phos	hemoglobin, p	rotein-	tyrosine	phosphat	case, ∠',	etical 3'-cuclic
			NT	<u>AA</u>	Score	Probability
ORF Name		AAID	Length	Length	211	3.8e-17
4554753_f2_45	1158 63	80	161 4	.86		
Protein name					s Name	Acc#
				sp:YQG	C_BACSU	P54486
Description						
HYPOTHETICAL 17.3	KD PROTEIN IN	$CCC\Delta - S$	TIVIT ACO	RGENIC	REGION	l
1111 0 1112 1 2 3 1 1	100 1110	- CCCA C				-
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
	NTID		<u>NT</u> Length	AA		Probability 4.7e-51
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	
ORF Name 4570341_f1_25	<u>NTID</u>	AAID	NT Length	<u>AA</u> Length	Score 531 s Name	4.7e-51
ORF Name 4570341_f1_25 Protein name	<u>NTID</u>	AAID	NT Length	AA Length 792 Locu	Score 531 s Name	4.7e-51 <u>Acc#</u>
ORF Name 4570341_t1_25  Protein name  putative glycosyl	NTID [1159 63 transferase.	AAID 81	NT Length	AA Length 792 Locu	Score 531 s Name	4.7e-51 <u>Acc#</u>
ORF Name  45.70.341_f1_25  Protein name  putative glycosyl  Description	NTID [1159 63 transferase.	AAID 81	NT Length	AA Length  792  Locu  gp:SC6	Score  531  s Name  D7	4.7e-51 <u>Acc#</u> AL133213
ORF Name  45.70.341_f1_25  Protein name  putative glycosyl  Description	NTID [1159 63 transferase.	AAID 81	NT Length 263	AA Length  Docu  Sp:SC6  AA Length	Score    531     s Name     D7     Score	Acc# AL133213  Probability
ORF Name  4570341_f1_25  Protein name  putative glycosyl  Description  Streptomyces coel	NTID  1159 63  transferase.  icolor cosmid  NTID	<u>AAID</u> 81 6D7.	NT Length 263	AA Length  792  Locu  gp:SC6	Score  531  s Name  D7	Acc# AL133213  Probability  7.4e-41
ORF Name  457.0341_f1_25  Protein name  putative glycosyl  Description  Streptomyces coel	NTID  1159 63  transferase.  icolor cosmid  NTID	AAID  6D7.  AAID	NT Length 263	AA Length  Sp:SC6  AA Length  AA Length  Locu  Locu  Locu	Score    531     S Name     D7     Score     439     Is Name	Acc# AL133213  Probability  7.4e-41  Acc#
ORF Name  457.0341_f1_25  Protein name  putative glycosyl  Description  Streptomyces coel  ORF Name  472.00_f1_13	NTID  1159 63  transferase.  icolor cosmid  NTID	AAID  6D7.  AAID	NT Length 263	AA Length  Sp:SC6  AA Length  AA Length  Locu  Locu  Locu	Score    531     S Name     D7     Score     439	Acc# AL133213  Probability  7.4e-41
ORF Name  457.0341_f1_25  Protein name  putative glycosyl  Description  Streptomyces coel  ORF Name  472.00_f1_13	NTID  1159 63  transferase.  icolor cosmid  NTID	AAID  6D7.  AAID	NT Length 263	AA Length  Sp:SC6  AA Length  AA Length  Locu  Locu  Locu	Score    531     S Name     D7     Score     439     Is Name	Acc# AL133213  Probability  7.4e-41  Acc#

ORF Name	NTID AAID	<u>NT</u> <u>AA</u> <u>Sco</u> <u>Length</u> <u>Length</u>	re Probability
4876090_c1_82	1161 6383	204 615 122	0.00012
Protein name		Locus Nam	e <u>Acc#</u>
		sp:MFD_BACS	P37474
Description			,
TRANSCRIPTION-R	EPAIR COUPLING FACTO	R (TRCF)	
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Sco Length Length	re Probability
4876300_c1_88	1162 6384	65 198	
Protein name		Locus Nam	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Sco Length Length	re Probability
ORF Name 7.15.7.5.7.6		Scc	Probability
		Length Length	1.1e-88
7.157.576f36.7 Protein name		Length         Length           642         1929           886	1.1e-88
7.157.576f36.7 Protein name	1163 [6385	Length Length Sco 642 1929 886 Locus Nar	1.1e-88 ne <u>Acc#</u>
Protein name  conserved hypot	1163 [6385	Length Length Score Locus Nat Dir: C72391	1.1e-88 ne Acc# C72391
Protein name  conserved hypot	1163 [6385	Length Length Score   1929   886   Locus Name   Print C72391   NT   AA   Score   Score   Score   Score   AA   Score   Score   Score   Score   AA   Score   Score   Score   Score   Score   Score   Score   AA   Score   Score	1.1e-88  ne Acc#  C72391  pre Probability
Protein name  conserved hypot  Description	ntid AAII	Length Length Score   1929   886   Locus Name   Print C72391   NT   AA   Score   Score   Score   Score   AA   Score   Score   Score   Score   AA   Score   Score   Score   Score   Score   Score   Score   AA   Score   Score	1.1e-88 ne Acc# C72391
Protein name  conserved hypot  Description  ORF Name	ntid AAII	Length Length Score   1929   886   Locus Name	1.1e-88  ne Acc#  C72391  pre Probability  1.9e-18  ne Acc#
Protein name  conserved hypot  Description  ORF Name  835175_f1_28	ntid AAII	Length Length Score   1929   886   Locus Name   Poir: C72391   NT	1.1e-88  ne Acc#  C72391  pre Probability  1.9e-18  ne Acc#
Protein name  conserved hypot  Description  ORF Name  835175_f1_28  Protein name	ntid AAII	Length Length Score   1929   886   Locus Nat   Pir:C72391	1.1e-88  ne Acc#  C72391  pre Probability  1.9e-18  ne Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
10437958_c3_133	1165	6387	135 4	.08		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA		_ 1 1 1 1 1 1
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
10647055 <u></u> £2 <u></u> 17	1166	6388	263	792	124	5.7e-06
Protein name				Locus	Name	Acc#
transcription regul	ator, crp	family		pir:F72	2285	F72285
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name	NTID 1167	<u>AAID</u>	Length		<u>Score</u> 876	Probability
			Length	Length		ŵ.
1406526 <u></u> c2 <u></u> 128			Length	Length 1254 Locus	876	1.3e-87
1406526 <u></u> c2 <u></u> 128			Length	Length 1254 Locus	876 s Name	1.3e-87 <u>Acc</u> #
1406526c2128 Protein name	1167		Length	Length 1254 Locus	876 s Name	1.3e-87 <u>Acc</u> #
Protein name  Description	1167		Length 417	Length  Locus  sp:PAT	876 S Name B_BACSU	1.3e-87 <u>Acc#</u> Q08432
Protein name  Description	1167		Length	Length 1254 Locus	876 s Name	1.3e-87 Acc# Q08432 Probability
Protein name  Description  PUTATIVE AMINOTRANS	1167  SFERASE B,  NTID	6389	Length 417  NT Length	Length Locus Sp:PAT	876 S Name B_BACSU	1.3e-87 <u>Acc#</u> Q08432
Protein name  Description  PUTATIVE AMINOTRANS  ORF Name	1167  SFERASE B,  NTID	AAID	Length 417  NT Length	Length Locus Sp:PAT	876 S Name B BACSU Score	1.3e-87 Acc# Q08432 Probability
Protein name  Description  PUTATIVE AMINOTRANS  ORF Name  14494530_f1_6	I167 SFERASE B, NTID	AAID 6390	Length 417  NT Length 862	Length Locus Sp:PAT	876 S Name B BACSU  Score  169 S Name	1.3e-87 Acc# Q08432 Probability 1.2e-08

	_,	NT AA Score Probability
ORF Name	NTID AAID	Length Length
14648577_c1_93	1169 6391	368 1107 174 1.1e-10
Protein name		Locus Name Acc#
transmembrane sens	sor	gp:AF051691 AF051691
Description		
Pseudomonas aerug transmembrane sens (fiuA) genes, comp	or (fiuR), and hydro	A (psfA), ECF sigma factor(fiuI),  examate-typeferrisiderophore receptor
ORF Name	NTID AAID	NT AA Score Probability
14875635_c3_139	1170 6392	313 942 602 1.4e-58
Protein name		Locus Name Acc#
conserved hypothe	tical protein ytqA	pir:D69999 D69999
Description		
		NT AA Gara Brobability
ORF Name	NTID AAID	Length Length Probability
2008893.7 <u></u> ±1 <u></u> 11	1171 6393	292 879 717 9.2e-71
Protein name		Locus Name Acc#
lipoic acid synth	ase	pir:A75480 A75480
Description		
		NT AA Score Probability
ORF Name	NTID AAID	Length Length
22478311 <u>c3</u> 145	1172 6394	149 450
Protein name		Locus Name Acc#
Description		*
NO-HIT		

	•		$\underline{\text{NT}}$	<u>AA</u>	Score	Probab	ility
ORF Name	NTID	AAID	Length	Length			
22708153_c3_132	1173	6395	364	.095	317	2.2e	-28
Protein name				Locus	s Name	<u>A</u>	.cc#
GIdB				gp:AF1	58372	AF	158372
Description							
Flavobacterium (gldB), GldC (gld hypothetical prot	dC) , and hypot	thetical	protein c	n gene, p genes, co	partial ompletec	cds;Gldf ds; and	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probab	ility
23620910_c2_111	1174	6396	85	258			
Protoin nama				Locu	s Name	<u> </u>	ACC#
<u>Protein name</u>							
Description					, ,		- 1
				:	1		
Description	NTID	AAID	NT Length	AA Length	Score	Probak	oility
Description		<u>AAID</u> 6397	Length		Score	Probak	
Description NO-HIT ORF Name			Length	Length		9.66	
Description NO-HIT ORF Name 24337765_t3_68	1175	6397	Length	Length	432 s Name	9.66	-68
Description NO-HIT  ORF Name  24337765_f3_68  Protein name	1175	6397	Length	Length 2712 Locu	432 s Name	9.66	e-68 Acc#
Description  NO-HIT  ORF Name  24337765t368  Protein name  115K outer membra	1175	6397	Length	Length 2712 Locu	432 s Name	9.66	e-68 Acc#
Description  NO-HIT  ORF Name  24337765f368  Protein name  115K outer members protein	1175	6397	Length	Length 2712 Locu	432 s Name	9.6€	e-68 Acc#
Description  NO-HIT  ORF Name  24337765t368  Protein name  115K outer member protein  Description	rane protein	6397	Length  904  SusC  NT Length	Length  2712  Locu  pir:JC	432 .s Name	9.66	e-68 Acc# C6027
Description  NO-HIT  ORF Name  24337765t368  Protein name  115K outer member protein  Description  ORF Name	rane protein	precursor	Length  904  SusC  NT Length	Length  Locu  pir:JC  AA  Length	432 s Name 6027 Score	9.66	e-68 Acc# C6027
Description  NO-HIT  ORF Name  24337765f368  Protein name  115K outer member protein  Description  ORF Name  24431510c3137	1175 rane protein  NTID  1176	precursor	Length  904  SusC  NT Length	Length  Locu  pir:JC  AA  Length	Score 108 Name	9.66 J Probal	e-68 Acc# C6027 Dility



ORF Name	NTID AAID	NT AA Score Probability
36360812_f2_36  Protein name	1182 6404	Locus Name Acc#
Description		
ORF Name	NTID AAID	NT AA Score Probability Length Length
3.640.6537f15	1183 6405	673 2022 344 2.8e-40
Protein name		Locus Name Acc#
		gp:SC9745
Description		-
S.cerevisiae chro	omosome XIII cosmid	9745.
		NIT 7 7
ORF Name	NTID AAID	NT AA Score Probability Length Length
ORF Name 3944711f352	-	— Score Probability
	-	Length Length Score Probability
3944711 <u></u> £3 <u></u> 52	1184 6406	Length Length Score Probability  207 624 399 4.6e-37
3944711_f3_52 Protein name	1184 6406	Length Length Score Probability  207 624 399 4.6e-37  Locus Name Acc#
3944711_f3_52 Protein name probable translat	1184 6406	Length         Length         Score         Probability           207         624         399         4.6e-37           Locus         Name         Acc#           pir:F64874         F64874
3944711_f3_52 Protein name probable translat	1184 6406	Length Length Score Probability  207 624 399 4.6e-37  Locus Name Acc#
Protein name probable translat Description	I184 6406  ion factor yciO  NTID AAID	Length Length Score Probability  207 624 399 4.6e-37  Locus Name Acc#  pir:F64874 F64874  NT AA Score Probability
Protein name  probable translate  Description  ORF Name	I184 6406  ion factor yciO  NTID AAID	Length Length Score Probability  207 624 399 4.6e-37  Locus Name Acc#  pir:F64874 F64874  NT AA Length Length Score Probability
Protein name  probable translate  Description  ORF Name  3946886_f3_53		Length         Length         Score         Probability           207         624         399         4.6e-37           Locus         Name         Acc#           pir:F64874         F64874           NT         AA         Score         Probability           Length         Length         7.3e-07
Protein name  probable translate  Description  ORF Name  3946886_f3_53_  Protein name		Length         Length         Score         Probability           207         624         399         4.6e-37           Locus         Name         Acc#           pir:F64874         F64874           NT         AA         Score         Probability           Length         Length         7.3e-07           Locus         Name         Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4007687 fl 10		6408		238	1937	4.8e-200
		0400			Name	Acc#
Protein name					Name	
DPP IV				gp:AB00	8194	AB008194
Description						
Porphyromonas ging	ivalis gene	for DPP	IV, comp	lete cds		E
			NT	AA		December 1 de la companya del companya del companya de la companya
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability
4119037_c3_138	1187	6409	281 8	46	134	9.7e-07
Protein name				Locus	s Name	Acc#
two-component resp	onse regula	tor		pir:B6	9655	B69655
lytT-involved					4	
Description						
Description			NUD	7.7		
Description ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
	·	<u>AAID</u>	Length		Score	Probability  2.6e-50
ORF Name	·		Length	Length		
ORF Name 496.0937c2109	1188		Length	Length	524 s Name	2.6e-50
ORF Name 4960937c2109	1188		Length	Length L209 Locus	524 s Name	2.6e-50 Acc#
ORF Name 496.0937c2109  Protein name hypothetical prote	1188		Length	Length L209 Locus	524 s Name	2.6e-50 Acc#
ORF Name  496.0937c2109  Protein name  hypothetical prote  Description	in b2710	6410	Length 402	Length  Locus  pir:B6	524 s Name	2.6e-50 Acc#
ORF Name  496.0937_c2_109  Protein name  hypothetical prote  Description  ORF Name	in b2710	AAID	Length  402  NT Length	Length  Locus  pir:B6  AA  Length	524 s Name 5051 Score	2.6e-50  Acc#  B65051  Probability
ORF Name  496.0937c2109  Protein name  hypothetical prote  Description	in b2710	6410	Length  402  NT Length	Length  Locus  pir:B6	524 s Name 5051	2.6e-50 Acc# B65051  Probability  1.3e-34
ORF Name  496.0937_c2_109  Protein name  hypothetical prote  Description  ORF Name	in b2710	AAID	Length  402  NT Length	Length  Locus  pir:B6  AA  Length	524 s Name 5051 Score	2.6e-50  Acc#  B65051  Probability
ORF Name  496.0937_c2_109  Protein name  hypothetical prote  Description  ORF Name  6.7.72836_t3_47	in b2710  NTID  1189	AAID 6411	Length  402  NT Length	Length  Locus  pir:B6  AA  Length	524 S Name 5051 Score 376 S Name	2.6e-50 Acc# B65051  Probability  1.3e-34

		•	N.L	AA	Caoro	Probability	
ORF Name	NTID	AAID	Length	Length	Score	PIODADITICY	
969812_c3_144	1190	6412	204	515	164	3.7e-12	
Protein name				Locus	s Name	Acc#	
RNA polymerase yhdM	ECF-type sigma	a factor .	homolog	pir:C69	9826	C69826	
Description							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
975405 <u>c2</u> 100	1191	6413	245	738	339	1.0e-30	]
Protein name				Locus	s Name	Acc#	
sam-dependent	methytransfera	se		pir:C7	2086	C72086	
Description		-					
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name		<u>AAID</u>	Length		Score 922	Probability	
			Length	Length			ě
10011662 <u></u> c1 <u></u> 20.9			Length	Length  1374  Locu	922	1.7e-92	
10011662 <u></u> c1 <u></u> 20.9			Length	Length  1374  Locu	922 s Name	1.7e-92	j.
Protein name  Description		6414	Length 458	Length  1374  Locu	922 s Name	1.7e-92	ē
Protein name  Description	1192	6414	Length 458	Length  1374  Locus  sp:PRI	922 s Name	1.7e-92 <u>Acc#</u>	]
Protein name  Description	1192	6414	Length 458	Length  1374  Locu	922 s Name	1.7e-92	
Protein name  Description  PRIMOSOMAL PRO	TEIN N' (REPLI	CATION FA	Length  458  ACTOR Y)  NT Length	Length  Locus  Sp:PRI	922 s Name A_BACSU	1.7e-92 <u>Acc#</u>	]
Protein name  Description  PRIMOSOMAL PRO	TEIN N' (REPLI	CATION FA	Length  458  ACTOR Y)  NT Length	Length  Locus  Sp:PRI  AA  Length  1533	922 s Name A_BACSU Score	1.7e-92 Acc# Probability	
Protein name  Description  PRIMOSOMAL PRO  ORF Name  10350927t3138	1192  OTEIN N' (REPLI  NTID  1193	CATION FA	Length  458  ACTOR Y)  NT Length	Length  Locus  Sp:PRI  AA  Length  1533	922 s Name A_BACSU  Score  172 s Name	1.7e-92 Acc#  Probability  5.8e-16	

ORF Name	<u>NT AA</u> <u>Score Probability</u> NTID AAID <u>Length</u> <u>Length</u>
11767812 f1_20	1194   6416   265   798   281   1.5e-24
Protein name	Locus Name Acc#
two-component re	sponse regulator pir:B69655 B69655
Description	
ORF Name	NT <u>AA</u> Score Probability NTID AAID Length Length
1204052 <u></u> £1 <u></u> 26	
Protein name	Locus Name Acc#
	sp:YGEK_ECOLI Q46791
Description	
HYPOTHETICAL TR	ANSCRIPTIONAL REGULATOR IN KDUI-LYSS INTERGENIC REGION
	NT AA Score Probability
ORF Name	NT AA Score Probability  NTID AAID Length Length
ORF Name 13729582_t1_24	NT AA Score Probability  NTID AAID Length Length
ORF Name	NTID AAID Length Length Score Probability  Length Length 128 2.4e-08  Locus Name Acc#
ORF Name  13.729582f124  Protein name	NTID AAID Length Length Score Probability  Length Length 128 2.4e-08  Locus Name Acc#
ORF Name  13.729582_t1_24  Protein name  hypothetical pr	NTID AAID Length Length Score Probability  1196 6418 195 588 128 2.4e-08  Locus Name Acc#  pir:C72325 C72325
ORF Name  13.729582_t1_24  Protein name  hypothetical pr	NTID AAID Length Length Score Probability  1196 6418 195 588 128 2.4e-08  Locus Name Acc#  pir:C72325 C72325
ORF Name  13.729582_f1_24  Protein name  hypothetical pr  Description	NTID         AAID         Length         Length         Score         Probability           1196         6418         195         588         128         2.4e-08           Locus         Name         Acc#           Detein         pir:C72325         C72325           NTID         AAID         Length         Length         Score         Probability           Length         Length         Length         Length         Probability
ORF Name  13.729582f124  Protein name  hypothetical pr  Description  ORF Name	NTID         AAID         Length         Length         Score         Probability
ORF Name  13.729582f124  Protein name  hypothetical pr  Description  ORF Name  13.773262f141	NTID         AAID         Length         Length         Score         Probability           1196         6418         195         588         128         2.4e-08           Locus         Name         Acc#           Detein         pir:C72325         C72325           NTID         AAID         Length         Length         Score         Probability           Length         Length         Length         Length         Probability

	<del></del>	NT AA Score	Probability
ORF Name	NTID AAID	Length Length	Probability
14252182_£1_45	1198 6420	212 639 222	2.6e-18
Protein name		Locus Name	Acc#
resolvase		pir:S38652	S38652
Description			
		NTITI 7.7)	
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length Score	Probability
14354025 <u></u> £1 <u></u> 1	1199   6421	447 1344	
Protein name		Locus Name	Acc#
Description			
мо-ніт			
		) TO 3.7	
ORF Name	NTID AAID	NT AA Length Length Score	Probability
14570301 <u>f266</u>	1200 6422	65 198	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
		2.7	
ORF Name	NTID AAID	NT AA Length Length Score	Probability
14928462 <u></u> c3 <u></u> 289	1201   6423	63 192	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
		NITT 7.7)	
ORF Name	NTID AAID	NT AA Score	Probability
16673406 <u></u> £2 <u></u> 95	1202 6424	1260	
Protein name	<del></del>	Locus Name	Acc#
Description			
NO-HIT			<del>,                                      </del>

16832885_cl_170	ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
hypothetical protein					296	1723	2.3e-177
Description   NTID   AAID   Length   Length	Protein name				Locus	Name	Acc#
ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         1204         6426         165         498	hypothetical protein	1			pir:JQ1	.020	JQ1020
ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         Locus         I65         498         Locus         Name         Acc#           Description         NO-HIT         NTID         AAID         NTID         AAID         Length         Length         Score         Probability           L97.35306_L2_97         I205         6427         65         198         63         0.045           Protein name         Locus         Name         Acc#           Sp:SRD2_CAEEL         Q21767           Description         SRD-2 PROTEIN           ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         NTID         AAID         Length         Length         Probability           Protein name         Locus         Name         Acc#           Protein name         Locus         Name         Acc#	Description						
1204   6426   165   498						Score	Probability
Description   NO-HIT	ORF Name	NTID	AAID	Length	Length		
Description   NO-HIT	19614050 <u></u> c1 <u></u> 162	1204	6426	165	.98		
NO-HIT	Protein name				Locus	s Name	Acc#
ORF Name         NTID         AAID         Length         Length         Score         Probability           19735306_F2_97         1205         6427         65         198         63         0.045           Protein name         Locus Name         Acc#           Description         SRD-2 PROTEIN         Q21767           ORF Name         NTID         AAID         Length         Length         Score         Probability           197734337_F3_126         1206         6428         64         195           Protein name         Locus Name         Acc#           Description	Description						
ORF Name         NTID         AAID         Length         Length         Score         Floadfilty           197.353.06t2_97	NO-HIT						
ORF Name         NTID         AAID         Length         Length         Score         Floadfilty           197.353.06t2_97			<u> </u>	NT	AA	Cacac	Probability
Protein name	ORF Name	NTID	AAID		Length	Score	Probability
Sp:SRD2_CAEEL   Q21767	19735306£297	1205	6427	65	198	63	0.045
Description           SRD-2 PROTEIN         NT         AA         Score         Probability           ORF Name         NTID         AAID         Length         Length         Probability           197.73437t3126	Protein name				Locus	s Name	Acc#
ORF Name         NTID         AAID         Length         Length         Score         Probability           197.73.43.7t3126         1206         6428         64         195           Protein name         Locus Name         Acc#           Description					sp:SRD	2_CAEEL	Q21767
ORF Name         NTID         AAID         Length         Length         Score         Probability           197.7343.7t3126	Description				- ,		
ORF Name         NTID         AAID         Length         Length         Score         Probability           197.73.43.7t3126         1206         6428         64         195           Protein name         Locus Name         Acc#           Description	SRD-2 PROTEIN					-	
ORF Name         NTID         AAID         Length         Length         Score         Probability           197.73.43.7t3126         1206         6428         64         195           Protein name         Locus Name         Acc#           Description				NT	7.7		
Protein name Locus Name Acc#  Description	ORF Name	NTID	AAID			Score	Probability
Description	19773437 <u></u> £3126	1206	6428	64	195		
	Protein name				Locu	s Name	Acc#
NO UTT	Description						
NO-HII	NO-HIT		a a		-		

ORF Name 20328267_c1_164	NTID	<u>AAID</u>	NT Length	AA Length	Score	Probability
Protein name		l ————————————————————————————————————	L L	Locus	Name	Acc#
Description						
NO-HIT						ī
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
20390875_£3_151	. 1208	6430	192	579	193	3.1e-15
Protein name				Locus	Name	Acc#
conserved hypothet:	ical prote	in		pir:E72	2312	E72312
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2043.76.75 <u></u> c2 <u></u> 241	. 1209	6431	235	708		
Protein name				Locus	Name	Acc#
Description						
NO-HIT					or .	
	w i		NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length	20020	220000222207
20.703426±294	. 1210	6432	67	204		
Protein name				Locus	Name	Acc#
Description						
ио-ніт						
NO-HIT			NT	AA		Durch all did to
NO-HIT ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
		<u>AAID</u> 6433	<u>Length</u>	*****	Score	Probability 3.3e-09
ORF Name			<u>Length</u>	Length		
ORF Name 20.723160c2218	1211	6433	<u>Length</u>	Length	138 Name	3.3e-09

ODE Name	NIMITIN AATIN	<u>NT</u> <u>AA</u> Length Length	Score	Probability
ORF Name	NTID AAID			
20976426_£3_114	1212 6434	91 276	124	3.9e-07
Protein name		Loci	ıs Name	Acc#
asparaginase homol	.og yccC	pir:F	9754	F69754
Description 7				
		NT AA	G	D11-1-1-1
ORF Name	NTID AAID	Length Length	Score	Probability
21647925 <u></u> ±1 <u></u> 7	1213 6435	459 1380	1231	3.1e-125
Protein name		Loci	ıs Name	Acc#
		sp:DC	JB_HAEIN	P44855
Description				
ANAEROBIC C4-DICAL	RBOXYLATE TRANSPORT	ER DCUB	(4)	- 8
		NT AA		
ORF Name	NTID AAID	Length Length	Score	Probability
21992175 <u></u> £3 <u></u> 117	1214 6436	327 984	84	0.0070
Protein name		Loci	ıs Name	Acc#
putative transmemb	orane efflux protei	n. gp:SC	791	AL132973
Description	. "			
Streptomyces coel:	icolor cosmid F91.	-		
		NTT 7.7.		
ORF Name	NTID AAID	NT AA Length Length	Score	Probability
22860128 <u></u> c2 <u></u> 219	1215 6437	83 252	64	0.031
Protein name		Loci	ıs Name	Acc#
		sp:SP	RC_XENLA	P36378
Description		L		
(OSTEONECTIN) (ON	) (BASEMENT MEMBRAN	E PROTEIN BM-40)		

ORF Name	<u>N'TID</u>	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
23617137_c3_268	1216	6438	224	75	267	4.5e-23
Protein name				Locus	s Name	Acc#
				sp:YJV	7_YEAST	P40893
Description						
HYPOTHETICAL 22.0	KD PROTEIN	IN HXT11	-HXT8 INT	ERGENIC	REGION	
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
23631252_c2_239	1217	6439	229 6	90		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						*
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 23.6.3.1.5.50		<u>AAID</u>	Length		Score	Probability 5.0e-08
			Length	Length		
23631550 <u></u> c1174	1218		Length	Length	125 Name	5.0e-08
23631550_c1_174 Protein name	1218		Length	Length 72 Locus	125 Name	5.0e-08 <u>Acc#</u>
23631550_cl174 Protein name  hypothetical prote	1218		Length  [123] [3]	Length  72  Locus  pir:A64	125 Name	5.0e-08 <u>Acc#</u>
23631550_cl174 Protein name  hypothetical prote	1218		Length	Length 72 Locus	125 Name	5.0e-08 <u>Acc#</u>
23631550_c1_174 Protein name  hypothetical prote  Description	1218 ein MJ1618 NTID	6440	Length  123  NT Length	Length Locus pir:A64	125 Name	5.0e-08 <u>Acc#</u> A64502
23631550_cl174 Protein name hypothetical prote Description  ORF Name	1218 ein MJ1618 NTID	6440	Length  123  NT Length	Length Locus  Locus  pir:A64  AA  Length	125 Name 1502 Score	5.0e-08  Acc#  A64502  Probability
Protein name  hypothetical protein  Description  ORF Name  23632787_c1_193	1218  ein MJ1618  NTID  1219	AAID 6441	Length  123  NT Length	Length Locus  Locus  pir:A64  AA  Length	125 Score 110 Name	5.0e-08  Acc#  A64502  Probability  3.6e-05

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24009637_c3_256		442		31	67	0.025
			<u> </u>	Locus	Name	Acc#
Protein name				sp:VE2		Q80910
				BP. VB2_		
Description						
REGULATORY PROTEIN	1 E2					
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24066056 £3_155		5443	91 2	276	134	5.5e-09
			L	Locus	Name	Acc#
Protein name				sp:CEBA		P23939
				DP . 02.2		
Description						
	CAN IN THE CO.					
BAMHI CONTROL ELE	MENT					
BAMHI CONTROL ELE	MENT	AAID	NT Length	<u>AA</u> Length	Score	Probability
	NTID	<u>AAID</u>	Length		Score	Probability
ORF Name 24236057t273	NTID		Length	Length	Score  Score	Probability  Acc#
ORF Name  24236.057f273  Protein name	NTID		Length	Length		
ORF Name  24236.057f273  Protein name  Description	NTID		Length	Length		
ORF Name  24236.057f273  Protein name	NTID		Length	Length		
ORF Name  24236.057f273  Protein name  Description	<u>NTID</u>	5444	Length 60	Length  183  Locus		
ORF Name  24236.057f273  Protein name  Description  NO-HIT  ORF Name	NTID 1222 [6	AAID	Length  60  NT Length	Length  Locus  AA  Length	3 Name	Acc#
ORF Name  24236.057f273  Protein name  Description  NO-HIT	NTID 1222 [6	5444	Length  60  NT Length	Length  Locus  AA  Length	Score	Acc# Probability
ORF Name  24236.057f273  Protein name  Description  NO-HIT  ORF Name	NTID 1222 [6	AAID	Length  60  NT Length	Length  Locus  AA  Length	3 Name	Acc#
ORF Name  24236.057f273  Protein name  Description  NO-HIT  ORF Name  24353391f140	NTID 1222 [6	AAID	Length  60  NT Length	Length  Locus  AA  Length	Score	Acc# Probability

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
24394017_f3_153  Protein name	1224	6446	63	Locus	3 Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24485932 <u></u> c3 <u></u> 279	1225	6447	848	2547	106	0.0070
Protein name				Locus	Name	Acc#
arylesterase	:			gp:AF04	14683	AF044683
Description						
Agrobacterium radio (dla) gene, partial						i i
dihydrolipoamide deh						and putative
					Score	Probability
dihydrolipoamide deh	ydrogenas	e(dlh) gei	NT Length	ial cds.		
dihydrolipoamide deh	NTID	e(dlh) gen	NT Length	AA Length		
ORF Name  2461567c3281	NTID	e(dlh) gen	NT Length	AA Length	Score	Probability
ORF Name  2461567c3281  Protein name	NTID	e(dlh) gen	NT Length	AA Length	Score	Probability
ORF Name  2461567c3281  Protein name  Description	NTID	e(dlh) gen	NT Length	AA Length  Locus	Score	Probability  Acc#
ORF Name  2461567c3281  Protein name  Description	NTID	e(dlh) gen	NT Length	AA Length	Score	Probability
ORF Name  2461567c3281  Protein name  Description  NO-HIT	NTID 1226	e(dlh) gen AAID 6448	NT Length  NT Length  NT Length	AA Length Locus	Score S Name	Probability  Acc#
ORF Name  2461567c3281  Protein name  Description  NO-HIT  ORF Name	NTID  NTID	AAID  AAID  AAID	NT Length  NT Length  NT Length	AA Length Locus  AA Length Locus  AA Length	Score S Name	Probability  Acc#
ORF Name  2461567c3281  Protein name  Description  NO-HIT  ORF Name  24640807t253	NTID  NTID	AAID  AAID  AAID	NT Length  NT Length  NT Length	AA Length Locus  AA Length Locus  AA Length	Score  Score	Probability  Acc#  Probability

ORF Name	NTID AAID	<u>NT AA</u> Length Length	Score	Probability
24882203_f1_29  Protein name	1228 6450	[62] [189] Loca	ıs Name	Acc#
Description		-		
NO-HIT				
ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score	Probability
24882932 <u></u> c1 <u>.</u> .175	1229 6451	178 537	348	1.2e-31
Protein name		Locu	s Name	Acc#
adaptive response r	regulatory protein	gp:AF(	47839	AF047839
Description				
Pseudoalteromonas s response regulatory			_	r andadaptive
ORF Name	NTID AAID	NT AA Length Length	Score	Probability
ORF Name  [2531640 c2 242	NTID <u>AAID</u>	Length Length	Score	Probability 4.4e-32
2531640 <u></u> c2 <u>.</u> 242	NTID <u>AAID</u>	Length Length	352	4.4e-32
		Length Length 200 603 Locu	352 s Name	
2531640c2242 Protein name		Length Length	352 s Name	4.4e-32
2531640c2242 Protein name unknown	1230 6452	Length Length  200 603  Locu	352 .s Name 06034	Acc# AF006034
2531640_c2_242 Protein name unknown Description Clostridium pasteur	1230 6452	Length Length  200 603  Locu	352 .s Name 06034 e (dhaT)	Acc# AF006034  gene, complete
2531640_c2_242 Protein name unknown Description Clostridium pasteur	1230 6452	Length Length  200 603  Locu  gp:AFC	352 .s Name 06034	Acc# AF006034
Protein name  unknown  Description  Clostridium pasteur cds.	1230 6452	Length Length  200 603  Locu  gp:AF0  ediol dehydrogenas	352 .s Name 06034 e (dhaT)	AF006034  Gene, complete  Probability
Protein name  unknown  Description  Clostridium pasteur cds.	ntid AAID	Length Length  200 603  Locu  gp:AFC  ediol dehydrogenas  NT AA  Length Length  78 237	352 .s Name 06034 e (dhaT)	AF006034  Gene, complete  Probability
Protein name  unknown  Description  Clostridium pasteur cds.  ORF Name	ntid AAID	Length Length  200 603  Locu  gp:AFC  ediol dehydrogenas  NT AA  Length Length  78 237	352 s Name 06034 e (dhaT)	Acc# AF006034  gene,complete  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
25914666 ±3_147		5454		79	86	0.010
Protein name	_			Locus	Name	Acc#
probable serine-th	reonine-prot	ein kina	ase	pir:T41	341	T41341
Description						
			NT	AA	Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	Length		
26178177 <u></u> c2 <u></u> 221	1233	6455	172	19	126	3.9e-08
Protein name				Locus	Name	Acc#
hypothetical prote	ein MTH847	-		pir:A6	9213	A69213
Description						
		-	NT	7.7		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 263.0663.729.7		<u>AAID</u> 6456	Length		Score	Probability  3.0e-24
			Length	Length		
26306637 <u></u> c3 <u></u> 297			Length	Length  1347  Locus	303	3.0e-24
26306637 <u></u> c3 <u></u> 297			Length	Length  1347  Locus	303 s Name	3.0e-24
26306637c3297 Protein name	1234		Length 448	Length  1347  Locus	303 s Name	3.0e-24
Protein name  Description	1234	6456	Length 448	Length  1347  Locus	303 S Name A_BACSU	3.0e-24 <u>Acc#</u>
Protein name  Description	1234	6456	Length  448	Length  Locus  sp:PRI	303 s Name	3.0e-24
Protein name  Description  PRIMOSOMAL PROTEI	N N/ (REPLIC	6456 ATION FA	Length  448  CTOR Y)  NT  Length	Length Locus Sp:PRI	303 S Name A_BACSU	3.0e-24 <u>Acc#</u>
Protein name  Description  PRIMOSOMAL PROTEI	N N/ (REPLIC	ATION FA	Length  448  CTOR Y)  NT  Length	Length  Locus  Sp:PRI  AA  Length	303 S Name A_BACSU	3.0e-24 <u>Acc#</u>
Protein name  Description  PRIMOSOMAL PROTEI  ORF Name  26360312_c3_290	N N/ (REPLIC	ATION FA	Length  448  CTOR Y)  NT  Length	Length  Locus  Sp:PRI  AA  Length	303 S Name A BACSU Score	Acc# Probability

ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length <u>S</u>	core Probability
26369028_f1_8	1236 6458	323 972 91	2.0e-91
Protein name		Locus N	ame Acc#
		sp:ASG2_E	COLI P00805
Description			
AMIDOHYDROLASE II)	(L-ASNASE II) (CO	LASPASE)	
ORF Name	NTID AAID	NT AA Length Length	core Probability
26460937_f2_102	1237 6459	392 1179 59	0 2.6e-57
Protein name		Locus N	ame Acc#
mannose-1-phosphate	e guanylyltransfera	pir:H7230	Н72303
Description			
ORF Name	NTID AAID	NT AA Length Length	core Probability
ORF Name 26593800c3267		S	
		Length Length	0 3.0e-56
26593800 <u></u> c3 <u></u> 267		Length Length St	0 3.0e-56 ame <u>Acc#</u>
26593800 <u></u> c3 <u></u> 267		Length Length Solution 197 594 58	0 3.0e-56 ame <u>Acc#</u>
26593800 <u></u> c3 <u></u> 267 Protein name		Length Length Solution 197 594 58	0 3.0e-56 ame <u>Acc#</u>
Protein name  Description		Length Length Solution 197   594   588   Locus No.   Sp:YJV8_Y	0 3.0e-56 ame <u>Acc#</u>
Protein name  Description		Length Length St.  197   594   58  Locus N.  sp:YJV8_Y.	0 3.0e-56 ame <u>Acc#</u>
26593800c3267  Protein name  Description  (EC 2.3.1)	NTID AAID	Length Length St.  197   594   58  Locus N.  sp:YJV8_Y.	0 3.0e-56  ame Acc#  EAST P40892  Core Probability
Protein name  Description  (EC 2.3.1)  ORF Name	NTID AAID	Length Length Solve Locus No.  Sp:YJV8_Y  NT AA Length Length Solve Length Length Length	0   3.0e-56   ame
26593800c3267  Protein name  Description  (EC 2.3.1)  ORF Name  26595192f122	NTID AAID 1239 6461	Length Length 58  Locus N  Sp:YJV8_Y  NT AA  Length Length 58  Locus N  Locus N  Locus N  Locus N  Locus N	0 3.0e-56  ame Acc#  EAST P40892  Core Probability  4 2.3e-51  ame Acc#

ORF Name	NTID AAID	NT AA Score	Probability
26601062_c3_259	1240 6462	61 186 Locus Name	Acc#
Protein name  Description		LOCUS Name	<u> 2007</u>
NO-HIT			
ORF Name	NTID AAID	NT AA Length Length Score	Probability
29328501 <u></u> ±3 <u></u> 131	1241 6463	Locus Name	Acc#
Protein name  Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Score	Probability
31723562c2248 Protein name	1242 6464	1158 Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Score Length Length	Probability
3320802_c3_294 Protein name	1243 6465	234 Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Length Length Score	Probability
33245255_c1_188 Protein name	1244 6466	261 786 103 Locus Name	0.014 Acc#
hypothetical pro	tein 2	pir:S49113	S49113
Description			

	NTID AAID Length Length	Probability
ORF Name		6.3e-12
34078300_f3_152	1245 6467 284 855 182	
Protein name	Locus Name	Acc#
integrase	gp:BFU75371	U75371
Description		
(tnpC), excisionase	is transposon Tn4555 TnpA (tnpA), integrase (xis), mobilization protein (mobA), and betete cds; and unknown genes.	e(int), TnpC ca-lactamase
	NT AA Score	Probability
ORF Name	NTID AAID Length Length	
34104127_f3_141	1246 6468 808 2427 274	3.0e-20
Protein name	Locus Name	Acc#
	sp:IRGA_VIBCH	P27772
Description	· ·	
IRON-REGULATED OU'	ER MEMBRANE VIRULENCE PROTEIN PRECURSOR	
IRON-REGULATED OUT	NT AA	Probability
ORF Name		Probability
	NT AA NTID AAID Length Length	Probability
ORF Name	NT AA NTID AAID Length Length	Probability  Acc#
ORF Name  3.426.0.911f3150  Protein name	NTID AAID Length Length Score 1247 6469 166 501	
ORF Name  3.426.0.911f3150  Protein name  Description	NTID AAID Length Length Score 1247 6469 166 501	
ORF Name  3.426.0.911f3150  Protein name	NTID AAID Length Length Score 1247 6469 166 501	
ORF Name  3.426.0.911f3150  Protein name  Description	NTID AAID Length Length Score 1247 6469 166 501	Acc#
ORF Name  3.426.0911f3150  Protein name  Description  NO-HIT	NTID AAID Length Length  Locus Name  NT AA Score  1247 6469 166 501  Locus Name  NTID AAID Length Length Score	Acc#
ORF Name  3.426.0911f3150  Protein name  Description  NO-HIT  ORF Name	NTID AAID Length Length  Locus Name  NT AA Score  1247 6469 166 501  Locus Name  NTID AAID Length Length Score	Acc#
ORF Name  3.426.0911f3150  Protein name  Description  NO-HIT  ORF Name  3.5314080f2_65	NTID AAID Length Length  Locus Name  NTID AAID Length Length  Locus Name  NTID AAID Length Length  NTID AAID Length Length  1248 6470 716 2151	Acc# Probability

		N(T) 7. 7.	
ORF Name	NTID AAID	NT AA Length Length Score	Probability
35704786_£2_92	1249 6471	92 279 326	2.5e-29
Protein name		Locus Name	Acc#
integrase IntN1		gp:BUU51917	U51917
Description			<del></del>
Bacteroides unifo	ormis insertion eleme	ent NBU1 fragment, integra	seIntN1 gene,
		NT AA	
ORF Name	NTID AAID	Length Length Score	Probability
3938817_c2_247	1250 6472	480 1443 1319	1.5e-134
Protein name		Locus Name	Acc#
aspartate ammonia	ı-lyase	gp:WSAJ2933	AJ002933
Description			
Wolinella succind	ogenes aspA, dcuA ger	nes and partial ansA gene.	. *
ORF Name	NTID AAID	NT AA Score Length Length	Probability
ORF Name 3.940.943_f2_74		— Score	Probability  2.8e-11
		Length Length Score	
3940.943_f2_74		Length Length Score	2.8e-11
3940943_f2_74		Length Length Score    355   1068   180     Locus Name	2.8e-11 Acc#
3940943_f2_74 Protein name AlgZ Description	[6473] [6473]	Length Length Score    355   1068   180     Locus Name	2.8e-11 Acc# U52431
3940943_f2_74 Protein name AlgZ Description	[6473] [6473]	Length Length Score  Locus Name  [gp:PAU52431]  sensor AlgZ (algZ) gene, co	2.8e-11 Acc# U52431
3940943_f2_74 Protein name AlgZ Description	[6473] [6473]	Length Length Score    355   1068   180     Locus Name     gp:PAU52431	2.8e-11 Acc# U52431
Protein name  AlgZ  Description  Pseudomonas aerug	ginosa AlgR-cognate s	Length Length  Score  Length Length  Score  Locus Name  gp:PAU52431  sensor AlgZ (algZ) gene,co	2.8e-11  Acc#  U52431  mplete cds.
Protein name  AlgZ  Description  Pseudomonas aerug  ORF Name	ginosa AlgR-cognate s	Length Length  Score  Length Length  Locus Name  gp:PAU52431  sensor AlgZ (algZ) gene,co  NT AA Length Length Score	2.8e-11  Acc#  U52431  mplete cds.  Probability
Protein name  Algz  Description  Pseudomonas aerus  ORF Name  4000953_f2_87	1251   6473     6473	Length Length  Score  Length Length  Locus Name  GP:PAU52431  Sensor AlgZ (algZ) gene,co  NT AA Length Length  Score  301 906 147	2.8e-11  Acc# U52431  mplete cds.  Probability  3.1e-08
Protein name  Algz  Description  Pseudomonas aerug  ORF Name  4000953_f2_87	1251   6473     6473	Length Length  Score  Length Length  Locus Name  Gp:PAU52431  Sensor AlgZ (algZ) gene,co  NT AA Length Length  Score  Locus Name  Locus Name	2.8e-11  Acc# U52431  mplete cds.  Probability  3.1e-08  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4072187_c1_172	1253	6475	587 1	.764	483	9.3e-95	
Protein name				Locus	Name	Acc#	
				sp:DXS	BACSU	P54523	
Description							
PROBABLE 1-DEOXYXYI	JULOSE-5-P	HOSPHATE	SYNTHASE	(DXP SYN	NTHASE)		
			NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length			
423162_c1_207	1254	6476	221	66			
Protein name				Locus	s Name	Acc#	
Description							
NO-HIT							
×.			NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length			_
ORF Name 434533.7±3110	<u>NTID</u>	<u>AAID</u> 6477	Length		Score	3.8e-09	
			Length	Length 1206 Locu	143 s Name	3.8e-09 <u>Acc#</u>	
434533.7 <u></u> £3 <u></u> 110			Length	Length	143 s Name	3.8e-09	
Protein name  Description	1255	6477	Length 401	Length  Locu  gp:ECA	143 s Name	3.8e-09 <u>Acc#</u> X02307	
4345337t3110 Protein name	1255	6477	Length 401	Length  Locu  gp:ECA	143 s Name	3.8e-09 <u>Acc#</u> X02307	
Protein name  Description	1255	6477	Length  401  partate a	Length Locu  gp:ECA  ammonia-	143 s Name	3.8e-09 <u>Acc#</u> X02307	
Protein name  Description	1255	ase (L-as	Length  401  partate a  NT  Length	Length  Locu  gp:ECA  ammonia-  AA  Length	143 s Name SPA lyase) (	3.8e-09  Acc#  X02307  EC4.3.1.1).	]
Protein name  Description  E. coli aspA gene	1255 for aspart	6477	Length  401  partate a  NT  Length	Length  Locu  gp:ECA  ammonia-  AA  Length	143 s Name SPA lyase) ( Score	3.8e-09  Acc#  X02307  EC4.3.1.1).  Probability  1.0e-08	
Protein name  Description  E. coli aspA gene  ORF Name	1255 for aspart	ase (L-as	Length  401  partate a  NT  Length	Length  Locu  gp:ECA  ammonia-  AA  Length  Locu  Locu	143 s Name SPA lyase) ( Score 171 s Name	3.8e-09  Acc#  X02307  EC4.3.1.1).  Probability  1.0e-08  Acc#	
Protein name  Description  E. coli aspA gene  ORF Name	1255 for aspart	ase (L-as	Length  401  partate a  NT  Length	Length  Locu  gp:ECA  ammonia-  AA  Length	143 s Name SPA lyase) ( Score 171 s Name	3.8e-09  Acc#  X02307  EC4.3.1.1).  Probability  1.0e-08	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4960312_±1_46	1257	5479	523 1	572	227	7.2e-16	]
Protein name	ν.			Locus	Name	Acc#	
putative integrase				gp:BA12	42593	AJ242593	
Description						.*	
Bacteriophage All8	complete ge	enome.					
			NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length			
579712_f1_42	1258	6480	339 1	.020			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA	Score	Probability	
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
	*	<u>AAID</u> 6481	Length		Score	Probability	
ORF Name	*		Length	Length	Score S Name	Probability  Acc#	
ORF Name 6.728452c3293	*		Length	Length	V		
ORF Name 6.7.28.452G3293  Protein name	*		Length	Length	V		
ORF Name 6.728452c3293  Protein name  Description	*		Length	Length	s Name	Acc#	
ORF Name 6.728452c3293  Protein name  Description	*		Length	Length 264 Locus	V		
ORF Name 6.728452_c3_293  Protein name  Description  NO-HIT	1259 NTID	6481	Length  87  NT Length	Length  Locus	s Name	Acc#	
ORF Name 6.728452_c3_293  Protein name Description NO-HIT ORF Name	1259 NTID	AAID	Length  87  NT Length	Length Locus  AA Length	Score	Acc# Probability	
ORF Name 6.728452_c3_293  Protein name Description NO-HIT  ORF Name  86.7688_t3_136	NTID	AAID	Length  87  NT Length	Length Locus  AA Length	Score  243  S Name	Acc# Probability 1.6e-20	

ř.
44
I,N
2220
123 420
13
FU
13
4.3
4
Has
200 AUG 400 AUG
13
1710
es una con
की दिए की संस्
or may

ΞY
<u> Y</u>
2
Ξ¥
912
ì
-
ΣY
5
5

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	<u>Probability</u>	
133301 £2 28	1265	6487	588	767	1406	9.0e-144	
Protein name			<u>L</u>	Locus	Name	Acc#	
				sp:SYQ_	ECOLI		
Description						~	
(GLNRS)	<del></del>						
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
1350033_f1_11	1266	6488	224	575			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA			-
ORF Name	NTID	AAID	Length	Length	Score	Probability	
1381287 <u></u> c2 <u></u> 91	1267	6489	172	519	470	1.4e-44	
Protein name				Locus	Name	Acc#	
				$sp:TPX_{\_}$	MYCTU	P95282	
Description							
PROBABLE THIOL PER	ROXIDASE,						
			NT	AA	Score	Probability	
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	BCOIC.		7
13859702 <u></u> £3 <u></u> 54	1268	6490	379	1140	181	3.8e-11	
Protein name				Locus	s Name	Acc#	
transposase	×			gp:AF0	38866	AF038866	
Description					- 2	-	-
Bacteroides fragi- protein BmpH (bmpH)	lis transpo genes, c	oson Tn552 omplete co	0 transp ls.	osase (b:	ipH) and	mobilization	

			NT	AA Langth	Score	Probability	
ORF Name	NTID	AAID	Length	Length			_
13945437_c3_98	1269	6491	155 4	68	86	0.00077	╛
Protein name				Locus	Name	Acc#	
				sp:DBH_	THEMA	P36206	
Description							
DNA-BINDING PROTEIN	1 HU						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
15738937_f1_24	1270	6492	104	315			
Protein name				Locus	Name	Acc#	
Description							
			-				
NO-HIT						**	
			NT	AA	Cacaro	Probability	-
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	-
	-	<u>AAID</u>	Length		Score	Probability	
ORF Name	-		Length	Length			
ORF Name 15909682_f3_52	1271	6493	Length	Length	137 S Name	1.9e-08	
ORF Name  1.5.90.96.82f352  Protein name	1271	6493	Length	Length  Locus	137 S Name	1.9e-08 <u>Acc</u> #	
ORF Name  15909682f352  Protein name  hypothetical protein	1271	6493	Length	Length 539  Locus pir:F70	137 S Name	1.9e-08 <u>Acc#</u> F70558	
ORF Name  15909682f352  Protein name  hypothetical protein	1271	6493	Length	Length  Locus	137 S Name	1.9e-08 <u>Acc</u> #	
ORF Name  15909682_f3_52  Protein name  hypothetical protein  Description	1271 in Rv1624c	6493	Length  212  NT  Length	Length Locus pir:F70	137 S Name	1.9e-08 <u>Acc#</u> F70558	
ORF Name  15909682_f3_52  Protein name  hypothetical protein  Description  ORF Name	1271 in Rv1624c	6493 <u>AAID</u>	Length  212  NT  Length	Length Locus  Locus  AA Length	137 S Name 0558 Score	1.9e-08  Acc#  F70558  Probability	
ORF Name  15909682f352  Protein name  hypothetical protein  Description  ORF Name  23475200t18	1271 in Rv1624c	AAID	Length  212  NT  Length	Length Locus  Locus  AA Length	137 S Name 0558 Score 179 S Name	1.9e-08  Acc#  F70558  Probability  1.1e-10	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24070786_c2_97	1273	6495	254	762	356	1.7e-32
Protein name				Locus	Name	Acc#
				sp:YQGI	_BACSU	P46339
Description						
REGION (ORF72)						
			NT	AA		D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
ORF Name	$\underline{\mathtt{NTID}}$	AAID	Length	Length	Score	Probability
26438887_c3_100	1274	6496	424	L275		
Protein name				Locus	Name	Acc#
Description						
NO-HIT	16					
NO-HII						
100-111			NT	AA		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
		AAID 6497	Length		Score	Probability
ORF Name			Length	Length	Score S Name	Probability  Acc#
ORF Name 29947188t115			Length	Length	7	
ORF Name 29947188f115  Protein name			Length	Length	7	
ORF Name  29947188f115  Protein name  Description			Length	Length  189  Locus	7	
ORF Name  29947188f115  Protein name  Description			Length	Length	7	
ORF Name  29947188_f1_15  Protein name  Description  NO-HIT	1275 NTID	6497	Length 62 NT Length	Length  Locus	s Name	Acc#
ORF Name  29947188_f1_15  Protein name  Description  NO-HIT  ORF Name	1275 NTID	AAID	Length 62 NT Length	Length  Locus  AA  Length	Score	Acc# Probability
ORF Name  29947188_f1_15  Protein name  Description  NO-HIT  ORF Name  30563966_f1_20	NTID	AAID	Length 62 NT Length	Length  Locus  AA  Length	Score 121 S Name	Acc# Probability 1.3e-07

	— ·	NT AA Score Probability
ORF Name	NTID AAID	Length Length
34179077_£1_16	1277 6499	262 789 170 1.2e-10
Protein name		Locus Name Acc#
		sp:EPSA_BURSO Q45407
Description		
EPS I POLYSACCHA	RIDE EXPORT OUTER MEM	BRANE PROTEIN EPSA PRECURSOR
		NT AA Goorg Probability
ORF Name	NTID AAID	Length Length Score Probability
36131937_c1_78	1278 6500	149 450 123 2.5e-07
Protein name		Locus Name Acc#
phosphate-bindin	g protein PstS	pir:H69097 H69097
Description		· **
		NT AA Grand Brobability
ORF Name	NTID AAID	Length Length Score Probability
4176337 <u></u> £3 <u></u> 55	1279 6501	470 1413 656 2.7e-64
Protein name		Locus Name Acc#
GumD protein		pir:S67820 S67820
Description		
		27
ORF Name	NTID AAID	NT AA Score Probability Length Length
4744002 <u></u> c2 <u></u> 89	1280 6502	198   597   249   3.6e-21
Protein name		Locus Name Acc#
hypothetical pro	otein (repA 5' region)	pir:S30120 S30120
Description		
ORF Name	NTID AAID	NT AA Score Probability Length Length
4867127£231	1281 6503	216 651 308 2.0e-27
Protein name		Locus Name Acc#
DedA family pro	tein	pir:B75253 B75253
Description		

ORF Name	NTID AAID	NT AA Length Length Score	Probability
6721890_c3_99	1282 6504	163 492 270	2.1e-23
Protein name		Locus Name	Acc#
N-acetylmuramoy.	l-L-alanine amidase ho	molog pir:G64126	G64126
Description			
ORF Name	NTID AAID	NT AA Score	Probability
7313162 <u></u> £1 <u></u> 4	1283   6505	285 858 410	3.1e-38
Protein name		Locus Name	Acc#
phosphate-binding	ng protein PstS	pir:H69097	Н69097
Description			
ORF Name	NTID AAID	NT AA Score	Probability
1212751 <u></u> c187	1284 6506	147 444	
Protein name		Locus Name	Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA Score	Probability
14103176c3114	<u> </u>	681 2046 246	2.5e-34
	1285 6507	2010	
Protein name	[1285 ] [6507	Locus Name	Acc#
	[1285] [6507]		Acc# U60208
		Locus Name	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
179762_c2_94	1286	6508	399	1200	465	9.2e-62	
Protein name				Locus	3 Name	Acc#	
				sp:YBD0	ECOLI	7	
Description						·	
HYPOTHETICAL 46.6	KD PROTEIN	IN PHEP-	NFNB INT	ERGENIC I	REGION		]
				7.7			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
22063387_c2_92	1287	6509	620	1863	378	9.5e-67	
Protein name				Locus	3 Name	Acc#	
alpha-1,3/4-fucos	idase precur	sor		gp:SSU	39394	U39394	
Description					*		
Streptomyces sp.	alpha-1,3/4-	fucosida	se precu	rsor gene	e, comple	etecds.	1
						i	l
			NT	7) 7)			}
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	-
ORF Name 238255_c3_113	<del></del>	<u>AAID</u>	Length		Score	Probability	
-	<del></del>		Length	Length		-	
238255 <u></u> c3113	<del></del>		Length	Length D69 Locus	106	0.0027	
238255 <u></u> c3113	<del></del>		Length	Length D69 Locus	106 Name	0.0027	
238255_c3_113 Protein name	1288	6510	Length	Length Locus	106 Name	0.0027	
Protein name  Description	1288	6510	Length  322	Length Locus Sp:YEH	106 Name	0.0027	
Protein name  Description	1288	6510	Length	Length Locus	106 Name	0.0027	
Protein name  Description  HYPOTHETICAL 27.9	1288  KD PROTEIN  NTID	IN MOLR-	Length  BGLX INTE	Length Locus sp:YEH	I06  Name LECOLI REGION	0.0027 <u>Acc#</u>	
Protein name  Description  HYPOTHETICAL 27.9  ORF Name	1288  KD PROTEIN  NTID	IN MOLR-	Length  BGLX INTE	Length Locus Sp:YEH  ERGENIC F  AA Length	I06  Name LECOLI REGION	0.0027 <u>Acc#</u>	
238255_c3_113  Protein name  Description  HYPOTHETICAL 27.9  ORF Name  24397305_c2_98	1288  KD PROTEIN  NTID	IN MOLR-	Length  BGLX INTE	Length Locus Sp:YEH  ERGENIC F  AA Length	106  Name CECOLI REGION Score	O.0027 Acc# Probability	
Protein name  Description  HYPOTHETICAL 27.9  ORF Name  24397305_c2_98  Protein name	1288  KD PROTEIN  NTID	IN MOLR-	Length  BGLX INTE	Length Locus Sp:YEH  ERGENIC F  AA Length	106  Name CECOLI REGION Score	Drobability  Acc#	

ORF Name	NTID AAID	NT AA Score Probability
24406550_c2_99	1290 6512	209 630 106 1.0e-05
Protein name		Locus Name Acc#
		gp:GGU25741 U25741
Description		
Group G streptococo	cus strain g6 emmL	gene, partial cds.
		NT AA
ORF Name	NTID AAID	Length Length Score Probability
25428436_c2_105	1291 6513	288 864
Protein name		Locus Name Acc#
Description		
NO-HIT		
ORF Name	NTID AAID	NT AA Score Probability
25897507 <u></u> £1 <u></u> .15	1292 6514	356 1071 105 0.038
Protein name		Locus Name Acc#
probable extracellu	lar nuclease	pir:D75625 D75625
Description		
ORF Name	NTID AAID	NT AA Score Probability
26210912 <u></u> 3 <u></u> 117	1293 6515	408 1227 109 0.0058
Protein name		Locus Name Acc#
silent surface laye	r protein	gp:AF079365 AF079365
Description		
Lactobacillus crisp		

			$\underline{\mathtt{NT}}$	$\underline{\underline{AA}}$	Score	Probability
ORF Name	NTID	AAID	Length	Length		
265878_f3_62	1294	6516	114 3	345	80	0.043
Protein name				Locus	Name	Acc#
MAR binding filamen	E-like pro	tein 1:MF	'P1	pir:T07	7111	T07111
protein				<u> </u>		<del></del> 3
Description						
			275	73.73		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2845427c189	1295	6517	939 2	2820	132	0.00012
Protein name				Locus	Name	Acc#
				sp:PFE	_PSEAE	Q05098
Description						
FERRIC ENTEROBACTIN	RECEPTOR	PRECURSOR	2		-	
<b>}</b>						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 3.01.920.86	<u>NTID</u>	<u>AAID</u> 6518	Length		Score 831	Probability 7.7e-83
* * * * * * * * * * * * * * * * * * * *			Length	Length		
3.01.920.86 <u></u> c1 <u></u> 7.6			Length	Length  L701  Locus	831	7.7e-83
30192086c176 Protein name			Length	Length  L701  Locus	831 Name	7.7e-83
30192086_c1_76Protein name  Description	1296	6518	Length	Length  L701  Locus	831 Name	7.7e-83
30192086c176 Protein name	1296	6518	Length 566	Length  L701  Locus	831 Name	7.7e-83
30192086_c1_76  Protein name  Description  BETA-GALACTOSIDASE,	1296 (LACTASE	6518	Length  566	Length Locus sp:BGA1	831 Name	7.7e-83
30192086c176  Protein name  Description  BETA-GALACTOSIDASE,  ORF Name	(LACTASE	6518 ) <u>AAID</u>	Length  566  NT  Length	Length  Locus  Sp:BGA1  AA  Length	831 S Name THEMA Score	7.7e-83  Acc#  Probability
30192086c176  Protein name  Description  BETA-GALACTOSIDASE,  ORF Name	1296 (LACTASE	6518	Length  566  NT  Length	Length Locus sp:BGA1	831 S Name THEMA	7.7e-83 <u>Acc#</u>
30192086c176  Protein name  Description  BETA-GALACTOSIDASE,  ORF Name	(LACTASE	6518 ) <u>AAID</u>	Length  566  NT  Length	Length  Locus  Sp:BGAI  AA  Length	831 S Name THEMA Score	7.7e-83  Acc#  Probability
Description  BETA-GALACTOSIDASE,  ORF Name  31256287_c3_116  Protein name  DNA-directed DNA po	(LACTASE  NTID  1297  lymerase,	AAID 6519  III chair	Length  NT Length  641	Length  Locus  Sp:BGAI  AA  Length	831 S Name Score 730 S Name	7.7e-83  Acc#  Probability  3.9e-72
Description  BETA-GALACTOSIDASE,  ORF Name  31256287c3116	(LACTASE  NTID  1297  lymerase,	AAID 6519  III chair	Length  NT Length  641	Length  Locus  Sp:BGAI  AA  Length  Locus  Locus	831 S Name Score 730 S Name	7.7e-83  Acc#  Probability  3.9e-72

ORF Name	NTID A	<u>N</u> AID Len	<u>T</u> <u>AA</u> gth Length	Score	Probability
34570437 f2 47				1142	8.5e-116
Protein name			Locu	s Name	Acc#
				D_ECOLI	P15288
Description					
(PEPTIDASE D)	*				
		N	T AA		
ORF Name	NTID A		gth Length	Score	Probability
35990807_c1_79	1299 652	1 222	669	611	1.6e-59
Protein name			Locu	s Name	Acc#
transaldolase-rela	ited protein		pir:G7	2394	G72394
Description	*				
		N	T AA		
ORF Name	NTID A	_	gth Length	Score	Probability
4101555 <u></u> c2 <u></u> 102	1300 652	2 332	999	144	2.0e-09
Protein name			Locu	s Name	Acc#
			gp:APU	72238	U72238
Description					
Anabaena PCC7120 (sequences.	ORFR1, ORFR2, C	RFR3, ORF	R4, and ORFR	5 genes,c	complete
/					
ORF Name	NTID A	_	<u>IT AA</u> agth <u>Length</u>	Score	Probability
447825 <u></u> c3 <u></u> 106	1301 652	502	1509	977	2.6e-98
Protein name			Locu	s Name	Acc#
			sp:BGA	L_BACME	052847
Description	•			•	
BETA-GALACTOSIDASI	E, (LACTASE)				×

· ·	-			73.73		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
4961691_f2_39	1302	6524	117	354		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length		
15112533 <u>f3</u> 19	. 1303	6525	355	1068	210	5.1e-15
Protein name				Locus	Name	Acc#
probable proteinase	PAB1960			pir:A7	5179	A75179
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
15820341 <u>f</u> 2 <u>14</u>	. 1304	6526	157	474	i	
Protein name				Locus	s Name	Acc#
Description						• ••
NO-HIT						
			NT	AA	*	
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
204811_f2_16	. 1305	6527	274	825	105	0.012
Protein name				Locus	s Name	Acc#
				gp:ATA	C012563	AC012563
Description						
Arabidopsis thalia	na chromos	ome I BAC	T23K23	genomic s	sequence	,complete
sequence.						

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
21664062_£1_5	1306	6528	168	507			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
26604687 <u></u> £2 <u></u> 12	1307	6529	205	518			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT			,				
							_
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 3.3.4.5.6.9.6.2		<u>AAID</u>	Length		Score	Probability	
·			Length	Length	Score Name	Probability  Acc#	
33456962 <u>£1</u> 2			Length	Length	. 5	-	
33456962f12			Length	Length	. 5	-	
Protein name  Description			Length	Length  382  Locus	. 5	-	
Protein name  Description			Length	Length	. 5	-	
Protein name  Description  NO-HIT	1308	6530	Length  293  NT  Length	Length  B82  Locus	: Name	Acc#	  ]
Protein name  Description  NO-HIT  ORF Name	NTID	AAID	Length  293  NT  Length	Length  Locus  AA  Length	Score	Acc# Probability	]
Protein name  Description  NO-HIT  ORF Name  4098437_f1_1	NTID	AAID	Length  293  NT  Length	Length  Locus  AA  Length	Score 115 Name	Acc# Probability 0.00062	]
Protein name  Description  NO-HIT  ORF Name  4098437_f1_1	NTID	AAID	Length  293  NT  Length	Length  Locus  AA  Length  Locus  Locus	Score 115 Name	Acc#  Probability  0.00062  Acc#	]

ORF Name NTID AAID Length Length Probabilit	Y
6837782_f1_4	
Protein name Locus Name Acc#	
Description	
NO-HIT	
NT AA Guun Duchahilih	
ORF Name NTID AAID Length Length Probabilit	Y
9939142 <u>t13</u> 1311	
Protein name Locus Name Acc#	
Description	
NO-HIT	
NT AA Grove Drobobilit	
ORF Name NTID AAID Length Length Probabilit	Y
10041563_c1_120 1312 6534 63 192	
Protein name Locus Name Acc#	
Description	
NO-HIT	
NT AA G	
ORF Name NTID AAID Length Length Score Probabilit	Y
10.726.552a3185 1313 6535 80 243	
Protein name Locus Name Acc#	
Description	
NO-HIT	
NT AA Grang Probabilit	
ORF Name NTID AAID Length Length Score Probabilit	ΣY
12500086 <u></u> c1119 1314	
Protein name Locus Name Acc#	
Description	
NO-HIT	

			NT	AA Tanath	Score	Probability
ORF Name	NTID	AAID	Length	Length	12.4.0	
13860653_c3_194	1315	6537	488	L467	1040	5.5e-105
Protein name				Locus	Name	Acc#
cell division prot	ein			gp:PAL:	249201	AJ249201
Description						
Prevotella albensi ORF-fts(partial).	s ftsQ (pa:	rtial), f	tsA and :	ttsZ gene	es and	
			NT	AA	Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	Length	<del></del>	*
14298312_c1_131	1316	6538	489	1470	1275	6.8e-130
Protein name				Locu	s Name	Acc#
				sp:MUR	C_PORGI	Q51831
Description						
ACETYLMURANOYL-L-A	LANINE SYN	THETASE)				10,
1						
			NT	AA		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 14644152_c2_164			Length		Score	Probability  6.4e-31
		AAID	Length	Length		
14644152 <u></u> c2164		AAID	Length	Length	341 s Name	6.4e-31
14644152 <u></u> c2 <u></u> 164		AAID	Length	Length 765 Locu	341 s Name	6.4e-31 <u>Acc#</u>
Protein name  FtsQ  Description	1317	<u>AAID</u> 6539	Length	Length  765  Locu  gp:AB0	341 s Name 04555	Acc# AB004555
14644152_c2_164 Protein name FtsQ	1317	<u>AAID</u> 6539	Length  254  [254]	Length Locu gp:AB0 , Fts2,	341 s Name 04555	Acc# AB004555  cds.
Protein name  FtsQ  Description	1317	<u>AAID</u> 6539	Length	Length  765  Locu  gp:AB0	341 s Name 04555	Acc# AB004555
Protein name  FtsQ  Description  Porphyromonas ging  ORF Name	jivalis gen	AAID 6539 es for Ft	Length  254  EsQ, FtsA  NT  Length	Length  Tocu  Gp:AB0  , FtsZ,	341 s Name 04555 complete	Acc# AB004555  cds.
Protein name  FtsQ  Description  Porphyromonas ging  ORF Name  164651C2169	jivalis gen	AAID 6539 es for Ft	Length  254  254  SQ, FtsA  NT  Length	Length  Cocu  Gp:AB0  FtsZ,  AA  Length	341 s Name 04555 complete Score	Acc# AB004555  cds.  Probability
Protein name  FtsQ  Description  Porphyromonas ging  ORF Name  164651_c2_169  Protein name	1317 givalis gen  NTID 1318	AAID 6539 es for Ft	Length  254  254  SQ, FtsA  NT  Length	Length  Locu  gp:AB0  , FtsZ,  AA  Length  2010  Locu	341 s Name 04555 complete Score 3334 s Name	Acc# AB004555  cds.  Probability  0.0
Protein name  FtsQ  Description  Porphyromonas ging  ORF Name  164651_c2_169  Protein name  DNA gyrase B subur	1317 givalis gen  NTID 1318	AAID 6539 es for Ft	Length  254  254  SQ, FtsA  NT  Length	Length  Cocu  Gp:AB0  FtsZ,  AA  Length	341 s Name 04555 complete Score 3334 s Name	Acc# AB004555  cds.  Probability  0.0  Acc#
Protein name  FtsQ  Description  Porphyromonas ging  ORF Name  164651_c2_169  Protein name	ivalis gen  NTID  1318	<u>AAID</u> 6539 es for Ft <u>AAID</u> 6540	Length  254  SQ, FtsA  NT  Length  669	Length  765  Locu  gp:AB0  , FtsZ,  AA  Length  2010  Locu  gp:AB0	341 s Name 04555 complete Score 3334 s Name 17713	Acc# AB004555  cds.  Probability  0.0  Acc# AB017713

			NT	7\ 7\		
ORF Name	NTID	AAID	<u>Length</u>	<u>AA</u> Length	Score	Probability
16593937_c1_127	1319	6541	435	L308	388	6.7e-36
Protein name				Locus	s Name	Acc#
				sp:YLA	O_BACSU	007639
Description						
HYPOTHETICAL 43.7	KD PROTEIN	IN NPRE-	PYCA INTI	ERGENIC P	REGION	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
16808437_c3_192	1320	6542	135	108	223	6.9e-18
Protein name				Locus	s Name	Acc#
UDP-N-acetylmuram	oylalanine-	D-glutama	te	pir:H7	0477	H70477
Description						
		<u> </u>	NT	ΑĀ		
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability
177253_c1_137	1321	6543	66	201	62	0.047
Protein name				Locu	s Name	Acc#
Orfs10c				gp:SCU	42227	U42227
Description						·
Saccharomyces cer subunit (MIP1) gen cds, and putative complete cds.	e, nuclear	gene enco	ding mit	ochondria	alprotei	n, partial
-			NT	AA		- 1121
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
19808211 <u></u> c1 <u></u> 135	1322	6544	181	546	289	2.1e-25
Protein name				Locu	s Name	Acc#
hypothetical prot	ein 1			pir:S7	0830	S70830
Description						

ODE Name	NULLD	7 7 TT	NT	<u>AA</u>	Score	Probability	
ORF Name	NTID	AAID	Length	Length			
20010316_£2_72	1323	6545	67	204			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT				1			
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
2149013_c3_187	1324	6546	343	L032	614	7.6e-60	
Protein name				Locus	Name	Acc#	
unknown				gp:EFU9	4707	U94707	
Description					-	*	
Enterococcus faecal yllB, yllC, yllD, pb complete cds.					-		
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
ORF Name 2151417t378	NTID 1325	<u>AAID</u>	Length		Score	Probability  2.5e-13	
		-	Length	Length			
2151417 <u>f3</u> 78		-	Length	Length	186 Name	2.5e-13	
2151417 <u>f3</u> 78		-	Length	Length Locus	186 Name	2.5e-13 <u>Acc#</u>	
2151417f378 Protein name	1325	6547	Length	Length  Locus  sp:YGY4	186 Name HALSQ	2.5e-13 <u>Acc#</u> P21562	]
Protein name  Description	1325	6547	Length  180  E	Length Locus Sp:YGY4 OF GYRA A	186 Name HALSQ	2.5e-13 <u>Acc#</u> P21562  (ORF 4)	
Protein name  Description	1325	6547	Length	Length Locus sp:YGY4 OF GYRA A	186 Name HALSQ	2.5e-13 <u>Acc#</u> P21562	
Protein name  Description  HYPOTHETICAL 80.2 K	1325 D PROTEIN	6547	Length  180  REGION C  NT  Length	Length Locus sp:YGY4 OF GYRA A	186 Name HALSQ	2.5e-13 <u>Acc#</u> P21562  (ORF 4)	
Protein name  Description  HYPOTHETICAL 80.2 K  ORF Name	D PROTEIN  NTID	IN THE 5	Length  180  REGION C  NT  Length	Length Locus Sp:YGY4  OF GYRA A  AA Length  329	186 Name HALSQ ND GYRB Score	2.5e-13  Acc# P21562  (ORF 4)  Probability	
2151417t378  Protein name  Description  HYPOTHETICAL 80.2 K  ORF Name  22303400c2166	D PROTEIN  NTID  1326	IN THE 5	Length  180  REGION C  NT  Length	Length Locus Sp:YGY4  OF GYRA A  AA Length  329	186 Name HALSQ ND GYRB Score 1291 Name	2.5e-13  Acc# P21562  (ORF 4)  Probability  1.4e-131	
Protein name  Description  HYPOTHETICAL 80.2 K  ORF Name  22303400_c2_166	D PROTEIN  NTID  1326	IN THE 5	Length  180  REGION C  NT  Length	Length Locus Sp:YGY4  AA Length  Jocus Locus	186 Name HALSQ ND GYRB Score 1291 Name	2.5e-13  Acc# P21562  (ORF 4)  Probability  1.4e-131  Acc#	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
23468837_£3_100	1327	6549	171	516		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
23.646.942 <u></u> c1126	1328	6550	328	987	368	8.9e-34
Protein name				Locus	Name	Acc#
				sp:MURI	BACSU	
Description						
ADDING ENZYME)		···				
ADDING ENZIME/					<del></del>	1
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
		<u>AAID</u>	Length		<u>Score</u>	Probability  3.9e-72
ORF Name		0	Length	Length	9	*
ORF Name 240.721.77c3175	1329	0	Length	Length	730 Name	3.9e-72
ORF Name  240.721.7	1329	0	Length	Length  Locus	730 Name	3.9e-72 <u>Acc#</u>
ORF Name  240.721.7	1329	0	Length	Length  Locus	730 Name	3.9e-72 <u>Acc#</u>
ORF Name  240.721.77c3175  Protein name  hypothetical prot  Description	ein NTID	6551	Length  452  NT  Length	Length  Locus  pir:S76	730 3 Name 5527	3.9e-72 Acc# S76527
ORF Name  240.721.77c3175  Protein name  hypothetical prot  Description  ORF Name	ein NTID	AAID	Length  452  NT  Length	Length  Locus  Locus  AA  Length	730 3 Name 5527	3.9e-72 Acc# S76527
ORF Name  240.721.77	ein NTID	AAID	Length  452  NT  Length	Length  Locus  Locus  AA  Length	730 Name 5527 Score	3.9e-72 Acc# S76527 Probability

$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	ty
24414077_f1_33   1331   6553   208   627	
Protein name Locus Name Acc#	
Description	
NO-HIT	
ORF Name NTID AAID Length Length Score Probability	ty
245643t375	
Protein name Locus Name Acc#	
conserved hypothetical protein pir:H75460 H7546	0
Description	
ORF Name NTID AAID Length Length Score Probability	ty
24658577t268	
Protein name Locus Name Acc#	•
Description	
ро-ніт	
	ty
25.83.21.6.1t241	
Protein name Locus Name Acc#	
Description	
NO-HIT	
NID AA	
ORF Name NTID AAID Length Length Score Probability	ty
2620187 <u>13</u> 96	
Protein name Locus Name Acc#	
Description	
ио-ніт	

ODE Maria	NMTD	7 7 TD	NT	AA Tangth	Score	Probability	
ORF Name	NTID	AAID	Length	Length			
26750336_£1_37	1336	6558	347	044	1218	7.5e-124	
Protein name				Locus	Name	Acc#	
hemolysin A				gp:PMU2	27587	U27587	
Description							
Prevotella melanin	ogenica her	molysin A	(phyA) g	ene, com	plete co	ls.	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
2931518_c3_189	1337	6559	486 1	461	892	2.6e-89	
Protein name				Locus	Name -	Acc#	
UDP-MurNac-tripept	ide synthet	ase	-	pir:E70	0450	E70450	
Description							
			· · · · · · · · · · · · · · · · · · ·				
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability	,
ORF Name 3016.643.7		<u>AAID</u>	Length		Score	Probability	
			Length	Length			T .
2016.6437_c3_191	1338	6560	Length	Length	75 Name	1.5e-06	Y .
Protein name  phospho-n-acetylmutransferase (mraY1)	1338	6560	Length	Length  49  Locus	75 Name	1.5e-06 <u>Acc#</u>	
2016.6437_c3_191	1338	6560	Length	Length  49  Locus	75 Name	1.5e-06 <u>Acc#</u>	
Protein name  phospho-n-acetylmutransferase (mraY1)	1338	6560	Length	Length  49  Locus  pir:E71	75 3 Name 664	1.5e-06 <u>Acc#</u> E71664	
Protein name  phospho-n-acetylmutransferase (mraY1)	1338	6560	Length  82  2	Length  49  Locus	75 Name	1.5e-06 <u>Acc#</u>	
Protein name  phospho-n-acetylmutransferase (mraY1)  Description	ramoy1-pent	6560 capeptide-	Length  82  2  NT Length	Length  49  Locus  pir:E71	75 3 Name 664	1.5e-06 <u>Acc#</u> E71664	
Protein name  phospho-n-acetylmutransferase (mraY1)  Description  ORF Name	ramoyl-pent RP595	Eapeptide-	Length  82  2  NT Length	Length  Locus  pir:E71  AA  Length	75 S Name 664 Score	1.5e-06  Acc#  E71664  Probability	
Protein name  phospho-n-acetylmutransferase (mraY1)  Description  ORF Name  31534452_f1_40	namoyl-pent RP595  NTID	AAID	Length  82  2  NT Length	Length  Locus  pir:E71  AA  Length	75 S Name Score 342 S Name	1.5e-06 Acc# E71664  Probability  1.6e-28	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
3166057_c2_163	1340	6562	389 [1	170	641	1.0e-62
Protein name			<u> </u>	Locus	Name	Acc#
PIOCEIII IIame				sp:MURG		
Description				P		
(EC 2.4.1)						
(EC 2.4.1.*)						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
33398557_f3_102	1341	6563	206	21		
Protein name				Locus	s Name	Acc#
Description					_	
NO-HIT						
	3.0		NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	1410	3.5e-39
33787927 <u>£1</u> 39	1342	6564	150	153	419	
Protein name					s Name	Acc#
				sp:DUT	_AQUAE	066592
Description						* * *
(DUTPASE) (DUTP P	YROPHOSPHAT	ASE)			-	
	16	-	NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	BCOLC	11000011101
33986038_f1_35	1343	6565	723	2172	164	9.8e-15
Protein name				Locu	s Name	Acc#
putative TonB-dep	endent oute	r membrar	ie	gp:AF0	48749	AF048749
Description				_		
Bacteroides fragi sequence.	lis capsula	ır polysa	ccharide .	biosynth	esis ope	ron, complete

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
34260912_c1_122	1344	6566	118	357	75	0.0099
Protein name				Locus	s Name	Acc#
hypothetical protei	n 2			pir:I4	759	
Description						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
3.7.542 <u>c2</u> 153	1345	6567	198	597	229	4.8e-19
<u>Protein name</u>				Locus	Name	Acc#
probable RNA polyme	rase sigma	factor		pir:T42	2015	T42015
Description						
			NT	ΔΔ		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 390927f111	NTID 1346	<u>AAID</u>	Length		Score	Probability  2.7e-14
		<u> </u>	Length	Length		
390927 <u></u> £1 <u></u> .11		<u> </u>	Length	Length	190 Name	2.7e-14
390927 <u></u> £1 <u></u> .11		<u> </u>	Length	Length  Locus	190 Name	2.7e-14 Acc#
390927_f1_11 Protein name	1346	6568	Length	Length  Locus	190 Name	2.7e-14 Acc#
Protein name  Description	1346	6568	Length	Length  Locus	190 Name 8868	2.7e-14  Acc#  AB028868
Protein name  Description	1346	6568	Length  168  5	Length  507  Locus  gp:AB02	190 Name	2.7e-14 Acc#
Protein name  Description  Mus musculus P4(21)	n mRNA, pa	6568	Length  168  5.  NT Length	Length  Locus  gp:AB02	190 Name 8868	2.7e-14  Acc#  AB028868
Protein name  Description  Mus musculus P4(21)  ORF Name	n mRNA, pa	e568  rtial cds	Length  168  5.  NT Length	Length  Locus  gp:AB02  AA  Length	190 Name 8868	2.7e-14  Acc#  AB028868
390927t111  Protein name  Description  Mus musculus P4(21)  ORF Name  3992135t395	n mRNA, pa	e568  rtial cds	Length  168  5.  NT Length	Length  Locus  gp:AB02  AA  Length	190 Name 8868 Score	2.7e-14  Acc#  AB028868  Probability

ORF Name	NTID AAID	NT AA Length Length Score	<u>Probability</u>
4079668_£2_66	1348 6570	414 1245 105	0.032
Protein name		Locus Name	Acc#
RING finger prote	ein	gp:AF036255	AF036255
Description			
Rattus norvegicus	RING finger protei	n mRNA, complete cds.	1.1
ORF Name	NTID AAID	<u>NT AA</u> Length Length Score	Probability
4174013_f1_7	1349 6571	265 798 101	6.7e-05
Protein name		Locus Name	Acc#
RecO		gp:HIU17037	U17037
Description			*
Haemophilus influ	enzae opacity assoc	iated proteins OapA and	Oan B. (care 2
oanB) genes compl	ete cds and DNA re	raced processis dapa and	Caps (caps and
		COMPINATION SHAROPSIK EK	
gene, partial cds.	cee eas, and DNA 16	combination andrepair pr	otein (recO)
gene, partial cds.	ecc cus, and biva re-	combination andrepair pr	otein (recO)
gene, partial cds.	ece das, and bha le		otein (recO)
gene, partial cds.  ORF Name	NTID AAID	NT AA Length Length Score	Probability
gene, partial cds.	NTID AAID	NT AA Score	
gene, partial cds.  ORF Name	NTID AAID	NT <u>AA</u> Score Length Length	Probability
ORF Name  48.75.812c3186  Protein name  DNA-binding prote	NTID AAID 1350 6572 in HB:DNA-binding pr	NT AA Score Length Length 84  Locus Name	Probability 0.0019
ORF Name  48.75.812c3186  Protein name  DNA-binding protein the protein protein the protein t	NTID AAID 1350 6572 in HB:DNA-binding pr	NT AA Score Length Length 84  Locus Name	Probability 0.0019
ORF Name  48.75.812c3186  Protein name  DNA-binding prote	NTID AAID 1350 6572 in HB:DNA-binding pr	NT AA Score Length Length 84  Locus Name	Probability 0.0019
ORF Name  48.75.812c3186  Protein name  DNA-binding protein the protein protein the protein t	NTID AAID 1350 6572 in HB:DNA-binding pr	NT AA Score Length Length 84  Locus Name Fotein pir:S00015	Probability 0.0019
ORF Name  48.75.812c3186  Protein name  DNA-binding protein the protein protein the protein t	NTID AAID 1350 6572 in HB:DNA-binding pr	NT AA Score Length Length 84  Locus Name	Probability 0.0019
ORF Name  48.75812c3186  Protein name  DNA-binding protein the discription	NTID AAID  1350 6572  in HB:DNA-binding process of the second seco	NT AA Score  Length Length Score  171 516 84  Locus Name  pir:S00015	Probability  0.0019  Acc#
ORF Name  48.75.812c3186  Protein name  DNA-binding protein the discription  Description  ORF Name	NTID AAID  1350 6572  in HB:DNA-binding process of the second seco	NT AA Score Length Length 84  Locus Name Fotein pir:S00015  NT AA Length Length Score	Probability  0.0019  Acc#  Probability
ORF Name  48.75.812c3186  Protein name  DNA-binding protein Description  ORF Name  49.57.837t265  Protein name	NTID AAID  1350 6572  in HB:DNA-binding process of the second seco	NT AA Score Length Length  171 516 84  Locus Name Fotein  Pir:S00015  NT AA Length Length Score  352 1059	Probability  0.0019  Acc#
ORF Name  48.75812c3186  Protein name  DNA-binding protein Description  ORF Name  495.7837t265  Protein name  Description	NTID AAID  1350 6572  in HB:DNA-binding process of the second seco	NT AA Score Length Length  171 516 84  Locus Name Fotein  Pir:S00015  NT AA Length Length Score  352 1059	Probability  O.0019  Acc#  Probability
ORF Name  48.75.812c3186  Protein name  DNA-binding protein Description  ORF Name  49.57.837t265  Protein name	NTID AAID  1350 6572  in HB:DNA-binding process of the second seco	NT AA Score Length Length  171 516 84  Locus Name Fotein  Pir:S00015  NT AA Length Length Score  352 1059	Probability  O.0019  Acc#  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
5117268_c1_121  Protein name	1352	6574	160	483 Locus	177 S Name	1.5e-13 Acc#	
				sp:YABE		P22186	
Description HYPOTHETICAL 17.4	KD PROTEIN	IN FRUR-1	FTSL INTE	ERGENIC R	EGION (	ORFC)	]
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
5994087_c1_123	1353	6575	708 2	2127	337	2.1e-34	
Protein name				Locus	Name	Acc#	
Description				sp:SP5D	BACSU	Q03524	
DIMETING DECIMALS			*				
BINDING PROTEIN)		*	<u>-</u> 1		vi		- (
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	- %
		<u>AAID</u> 6576	Length		Score	Probability	
ORF Name			Length	<u>Length</u>	344		
ORF Name 6.0.7.26.83			Length	Length	344 Name	8.1e-51	
ORF Name 60.7.26.83c1125  Protein name	1354	6576	Length 376	Length  131  Locus	344 Name	8.1e-51 <u>Acc#</u>	
ORF Name 60.7.26.83c1125  Protein name  Description	1354	6576	Length 376	Length  131  Locus	Name BORBU	8.1e-51 <u>Acc#</u> Q44776	
ORF Name 60.7.26.83c1125  Protein name  Description	1354	6576	Length  376  [1]	Length  131  Locus  Sp:MRAY	344 Name	8.1e-51 <u>Acc#</u>	
ORF Name 6.0.7.26.83c1125  Protein name  Description (UDP-MURNAC-PENTAPE	1354 EPTIDE PHOSI	PHOTRANSF	Length  376  ERASE)  NT Length	Length  Locus  Sp:MRAY  AA  Length	Name BORBU	8.1e-51 <u>Acc#</u> Q44776	
ORF Name 6.0.7.26.83c1125  Protein name  Description  (UDP-MURNAC-PENTAPE  ORF Name	1354 EPTIDE PHOSI	PHOTRANSF	Length  376  ERASE)  NT Length	Length  Locus  Sp:MRAY  AA  Length	Name BORBU  Score	8.1e-51 Acc# Q44776  Probability	
ORF Name 6.0.7.26.83	DTIDE PHOSI	PHOTRANSF  AAID	Length  376  ERASE)  NT Length	Length  Locus  Sp:MRAY  AA  Length	Name BORBU  Score  153  Name	8.1e-51 Acc# Q44776  Probability  5.4e-11	

ORF Name	NTID AA	ID Length	AA Score	Probability
6758437 f1 36	1356   6578		361   186	7.0e-13
Protoin name				
Protein name			Locus Name	Acc#
probable sulfolip	id biosynthesis	protein SqdA	pir:A42380	A42380
Description				
ORF Name	NTID AA	ID Length	<u>AA</u> <u>Length</u> <u>Score</u>	Probability
10328140 <u>f1</u> 3	1357 6579	71 2	16	
Protein name		-%-	Locus Name	Acc#
Description				
NO-HIT			*	
ORF Name	NTID AA	ID Longth	AA Score	Probability
	NIID AA.	ID <u>Length</u>	Length -	
10442793 <u></u> ±3 <u></u> 46			<u>Length</u>	, , , , , , , , , , , , , , , , , , ,
				<u>Acc#</u>
10442793 <u></u> £3 <u></u> 46			48	<u>Acc#</u>
10442793±346 Protein name			48	<u>Acc#</u>
Protein name  Description			48	) <u>Acc#</u>
Protein name  Description		115 3	48	Acc# Probability
Protein name  Description  NO-HIT	1358   6580	I15 3  NT Length	Locus Name  AA Score	
Protein name  Description  NO-HIT  ORF Name	1358   6580	I15 3  NT Length	Locus Name  Locus Name  AA Length Score	Probability
10442793_f3_46  Protein name  Description  NO-HIT  ORF Name  10744192_c1_95	NTID AA	III5 3  NT Length 281 8	Locus Name  AA Length  Score  153	Probability  8.1e-08
Protein name  Description  NO-HIT  ORF Name  10744192_c1_95.  Protein name	NTID AA	III5 3  NT Length 281 8	Locus Name  AA Score Length 153  Locus Name	Probability  8.1e-08  Acc#
Protein name  Description  NO-HIT  ORF Name  10.744192_c1_95  Protein name  potassium channel	NTID AAI 1359 6581 alpha subunit Kv	115 3  NT Length 281 8	Locus Name  AA Score Length  Locus Name  gp:XLU20342	Probability  8.1e-08  Acc#  U20342

		2.7	
ORF Name	NTID AAII	$rac{ ext{NT}}{ ext{Length}} rac{ ext{AA}}{ ext{Length}}$	Probability
12518961_c2_100	1360 6582	325 978 241	5.2e-28
Protein name		Locus Name	Acc#
probable protopo RP847	orphyrinogen oxidase	e (hemK) pir:G71646	G71646
Description		<del>,,,,,,</del>	
ORF Name	NTID AAII	NT AA Score Length Length	Probability
14256430 <u>f2</u> 37	1361 6583	191 576 445	6.1e-42
Protein name		Locus Name	Acc#
conserved hypoth	etical protein MTH7	00 pir:E69193	E69193
Description			
	-	NT AA	
ORF Name	NTID AAII	- Score	Probability
14508510 <u></u> £3 <u></u> 76	1362   6584	[62][186]	
		1 1 11 1	
Protein name		Locus Name	Acc#
Protein name  Description			Acc#
			Acc#
Description		Locus Name	Acc#
Description	NTID AAID	Locus Name  NT AA Score	
Description NO-HIT		Locus Name  NT AA Score	
Description NO-HIT ORF Name		Locus Name  NT AA Length Length	Probability
Description NO-HIT ORF Name 15835436c3152	[1363   6585	Locus Name  NT AA Length Length  308 927 228	Probability  6.1e-19
Description NO-HIT  ORF Name 15835436_c3_152  Protein name	[1363   6585	Locus Name  NT AA Length Length Score  308 927 228  Locus Name	Probability  6.1e-19  Acc#
Description  NO-HIT  ORF Name  15835436_c3_152  Protein name  hypothetical pro	[1363   6585	Locus Name  NT AA Score Length Length  228 Locus Name  pir:E69840	Probability  6.1e-19  Acc#
Description  NO-HIT  ORF Name  15835436_c3_152  Protein name  hypothetical pro	[1363   6585	Locus Name  NT AA Score  Length Length  228  Locus Name  pir:E69840	Probability  6.1e-19  Acc#
Description  NO-HIT  ORF Name  15835436_c3_152  Protein name  hypothetical pro  Description	tein yitL  NTID AAID	Locus Name  NT AA Score  Length Length  Locus Name  Locus Name  pir:E69840	Probability  6.1e-19  Acc#  E69840
Description  NO-HIT  ORF Name  15835436c3152  Protein name  hypothetical pro  Description  ORF Name	tein yitL  NTID AAID	Locus Name  NT AA Score Length Length  228 Locus Name pir:E69840  NT AA Length Length	Probability  6.1e-19  Acc#  E69840
Description  NO-HIT  ORF Name  15835436c3152  Protein name  hypothetical pro  Description  ORF Name  16892517c3134	tein yitL  NTID AAID	Locus Name  NT AA Score Length Length  Locus Name  Locus Name  pir:E69840  NT AA Score Length Length  70 213	Probability  6.1e-19  Acc#  E69840  Probability
Description  NO-HIT  ORF Name  15835436c3152  Protein name  hypothetical pro  Description  ORF Name  16892517c3134  Protein name	tein yitL  NTID AAID	Locus Name  NT AA Score Length Length  Locus Name  Locus Name  pir:E69840  NT AA Score Length Length  70 213	Probability  6.1e-19  Acc#  E69840  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
19953510_c2_102 Protein name	1365	6587	457	Locus	626 Name	4.1e-61 Acc#	
argininosuccinate  Description	lyase			pir:D70	0419	D70419	
			NT	AA	Score	Drobabilit	
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
236.73455 <u></u> c1 <u></u> 80	1366	6588	161	186	126	3.9e-08	
Protein name			٠	Locus	Name	Acc#	
				sp:RECX	PSEAE	P37860	
Description							
REGULATORY PROTEI	N RECX						
						······································	
			NT	7.7.			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 24026562		<u>AAID</u>	Length		Score 567	Probability 7.2e-55	
*			Length	Length			
24026562 <u></u> £3 <u></u> 59			Length	Length	567 Name	7.2e-55	
24026562 <u></u> £3 <u></u> 59			Length	Length  248  Locus	567 Name	7.2e-55 <u>Acc#</u>	_
24026562f359 Protein name			Length	Length  248  Locus	567 Name	7.2e-55 <u>Acc#</u>	
Protein name  Description			Length 415	Length  248  Locus  sp:ASSY	567 Name	7.2e-55 <u>Acc#</u>	
Protein name  Description			Length	Length  248  Locus	567 Name	7.2e-55 <u>Acc#</u>	
24026562_f3_59Protein name  Description  LIGASE)	1367 NTID	6589	Length  415 1  NT Length	Length  Locus  Sp:ASSY	567 Name _METJA	7.2e-55 <u>Acc#</u> Q60174	
24026562_f3_59Protein name  Description  LIGASE)  ORF Name	1367 NTID	AAID	Length  415 1  NT Length	Length  Locus  Sp:ASSY  AA  Length	Name METJA Score	7.2e-55 <u>Acc#</u> Q60174	
24026562_f3_59  Protein name  Description  LIGASE)  ORF Name  24300018_c3_133	1367 NTID	AAID	Length  415 1  NT Length	Length  Locus  Sp:ASSY  AA  Length  01	Name METJA Score	7.2e-55 Acc# Q60174 Probability	
Protein name  Description  LIGASE)  ORF Name  24300018_c3_133  Protein name	1367 NTID	AAID	Length  415 1  NT Length	Length  Locus  Sp:ASSY  AA  Length  01	Name METJA Score	7.2e-55 Acc# Q60174 Probability	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24353376_f1_6	1369	6591	533	1602	1279	2.6e-130
Protein name				Locus	s Name	Acc#
				gp:AB02	24946	AB024946
Description				<del>-</del>		*
Escherichia coli p	lasmid pB17	1 DNA, co	omplete s	sequence.	•	
			) TITI	73.73		
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
25663952_f1_2	1370	6592	188	67	280	1.9e-24
Protein name				Locus	s Name	Acc#
				sp:MTG/	A_ACICA	024849
Description				-		
(EC 2.4.2) (MONO)	FUNCTIONAL	TGASE)				
(EC 2.4.2) (MONO)	FUNCTIONAL	TGASE)				
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
	NTID		<u>Length</u>		Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	Length	Score S Name	Probability  Acc#
ORF Name  259.7518.7c3153	NTID	AAID	<u>Length</u>	Length		
ORF Name  25975187c3153  Protein name	NTID	AAID	<u>Length</u>	Length		
ORF Name  25975187_c3_153  Protein name  Description	NTID	AAID	Length  [221]	Length  Locus		
ORF Name  25975187_c3_153  Protein name  Description	NTID	AAID	<u>Length</u>	Length		
ORF Name  25975187c3153  Protein name  Description  NO-HIT	NTID  1371  NTID	<u>AAID</u> 6593	Length  221  NT Length	Length Locus	s Name	Acc#
ORF Name  25975187c3153  Protein name  Description  NO-HIT  ORF Name	NTID  1371  NTID	AAID 6593 AAID	Length  221  NT Length	Length Locus  AA Length	Score	Acc# Probability
ORF Name  25975187c3153  Protein name  Description  NO-HIT  ORF Name  26369087f2_44	NTID  1371  NTID  1372	<u>AAID</u> 6593 <u>AAID</u> 6594	Length  221  NT Length	Length Locus  AA Length	Score  554  Name	Acc# Probability  1.7e-53

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
32228388_c2_122	1373	6595	252	759	113	0.0016	
Protein name				Locus	Name	Acc#	
				sp:HEXA	A_BLADI	Q17127	
Description							
HEXAMERIN PRECURS	OR						
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
33367175_c3_131	1374	6596	163 4	.92			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
						×	
			ATT.	70.70			-
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
		<u>AAID</u> 6597	<u>Length</u>		Score	Probability 7.5e-53	]
ORF Name			<u>Length</u>	Length 81			]
ORF Name  3.6.211443f235	1375	6597	Length 326	Length 81	548 Name	7.5e-53	] .
ORF Name  3.6.211443f235  Protein name	1375	6597	Length 326	Length 81 Locus	548 Name	7.5e-53 Acc#	
ORF Name  3.6.211443f235  Protein name  N-acetyl-gamma-glu  Description	1375	6597 hate redu	Length  326  9  ctase,	Length  Locus  pir:F69	548 Name	7.5e-53 Acc# F69508	] '.
ORF Name  3.6.211443f235  Protein name  N-acetyl-gamma-glu	1375	6597	Length  326  9  ctase,	Length 81 Locus	548 Name	7.5e-53 Acc#	
ORF Name  3.6.211443f235  Protein name  N-acetyl-gamma-glu  Description	1375  atamy1-phosp.  NTID	6597 hate redu	Length  326  9  Ctase,  NT  Length	Length  Locus  pir:F69	548 Name	7.5e-53 Acc# F69508	] .
ORF Name  3.6.211443f235  Protein name  N-acetyl-gamma-glu  Description  ORF Name	1375  atamy1-phosp.  NTID	hate redu	Length  326  9  Ctase,  NT  Length	Length  Locus  pir:F69  AA  Length	548 Name	7.5e-53 Acc# F69508	
ORF Name  3.6.211.443	1375  atamy1-phosp.  NTID	hate redu	Length  326  9  Ctase,  NT  Length	Length  Locus  pir:F69  AA  Length	548 Name 5508	7.5e-53  Acc# F69508  Probability	]

ORF Name	NTID	AAID I	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4095050_f3_61	1377 6	599	260 7	83	375	1.6e-34	
Protein name					Name	<u>Acc#</u>	
pyrroline-5-carbo	exyrate reducta	ase ————		gp:CSAJ	10739	AJ010739	
Description							
Clostridium stic	clandii proC g	ene and 5	flank	ing regi	on.		
-			NT	AA	Score	Probability	
ORF Name	NTID	AAID I	Length	Length			
4377005_c3_130	1378 60	600 2	230 6	93	541	4.1e-52	
Protein name				Locus	Name	Acc#	
				sp:PYRE	BACSU	P25972	
Description					:	<u></u> ,	
OROTATE PHOSPHORI	BOSYLTRANSFER	ASE, (OPR	T) (OPR'	TASE)			
OROTATE PHOSPHORI	BOSYLTRANSFER	ASE, (OPR					
ORF Name	BOSYLTRANSFER		NT Length	AA Length	Score	Probability	
	NTID	AAID I	<u>NT</u> Length	AA	Score	Probability	
ORF Name	NTID	AAID I	<u>NT</u> Length	AA Length		Probability  Acc#	
ORF Name 4801552_t2_45  Protein name	NTID	AAID I	<u>NT</u> Length	AA Length			
ORF Name 4801552_t2_45  Protein name  Description	NTID	AAID I	<u>NT</u> Length	AA Length			
ORF Name 4801552_t2_45  Protein name	NTID	AAID I	<u>NT</u> Length	AA Length			
ORF Name  4801552_f2_45  Protein name  Description  NO-HIT	NTID	<u>AAID I</u>	NT Length	AA Length 377 Locus			
ORF Name  48.015.52f245  Protein name  Description  NO-HIT  ORF Name	NTID   66	AAID I	NT Length  NT Length	AA Locus  AA Length	Name	Acc#	
ORF Name  48.015.52f245  Protein name  Description  NO-HIT  ORF Name  48.040.51f234	NTID   66	AAID I	NT Length	AA Locus  AA Length	Name	Acc#	
ORF Name  48.015.52f245  Protein name  Description  NO-HIT  ORF Name	NTID   66	AAID I	NT Length  NT Length	AA Locus  AA Length	Name Score	Acc#	
ORF Name  48.015.52f245  Protein name  Description  NO-HIT  ORF Name  48.040.51f234	NTID   66	AAID I	NT Length  NT Length	AA Length  AA Length	Name Score	Acc# Probability	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4804813 f2 36		6603		1134	670	8.8e-66	
Protein name			L	Locus	Name	Acc#	
					BACSU	P36839	
Description							
ACETYLORNITHINE	AMINOTRANSFER	ASE, (ACC	DAT)				٦
						2	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
5110712_c2_114	1382	6604	413	1242	398	1.3e-37	
Protein name				Locus	s Name	Acc#	
sensory transdu slr2104:protein			:	pir:S75	5136	S75136	
Description				J			
-							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 5131925_c1_81		<u>AAID</u> 6605	Length		Score	Probability	7
-			Length	Length	*		
5131925c181	1383	6605	Length	Length	108 Name	0.033	
5131925_c1_81	1383	6605	Length	Length Locus	108 Name	0.033 <u>Acc#</u>	, , , , , , , , , , , , , , , , , , ,
5131925_cl_81	1383	6605	Length	Length Locus	108 S Name	0.033 <u>Acc#</u> T04772	, , , , , , , , , , , , , , , , , , ,
5131925_cl_81	1383	6605	Length	Length  Locus  pir:T04	108 Name	0.033 <u>Acc#</u>	
5131925_cl_81		0	Length 659  NT Length	Length Locus pir:T04	108 S Name	0.033 <u>Acc#</u> T04772	
Protein name  hypothetical pr  Description  ORF Name		6605 0	Length 659  NT Length	Length  Locus  pir:T04  AA  Length	108 Name 1772 Score	D.033  Acc#  T04772  Probability	
5131925_cl_81		6605 0	Length 659  NT Length	Length  Locus  pir:T04  AA  Length	108 Name 1772 Score 254 Name	0.033 Acc# T04772 Probability 1.1e-21	
Protein name  hypothetical pr  Description  ORF Name  5270050_f2_33  Protein name		6605 0	Length 659  NT Length	Length  Locus  Pir:T04  AA  Length  Locus  Locus	108 Name 1772 Score 254 Name	0.033   Acc#   T04772   Probability   1.1e-21   Acc#	

			$\underline{ ext{NT}}$	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length			
5270302_£1_11	1385 6	607	554 1	-665	1248	5.0e-127	
Protein name				Locus	s Name	Acc#	
acetyl-CoA syntheta	se related j	protein		pir:F6	9193	F69193	
Description							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
994002 <u></u> £3 <u></u> 53	1386	608	334 1	.005	273	1.0e-23	
Protein name				Locus	Name	Acc#	
probable malate deh dehydrogenase:protei				pir:S7	5735	S75735	
dehydrogenase:protei		ne coacit					
Description							
			NT	AA			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 10345327t2103		<u>AAID</u>	Length		Score	Probability 7.8e-40	
		-	Length	Length			
10345327t2103  Protein name  115K outer membrane	1387 66	609	Length	Length	410 Name	7.8e-40	]
Protein name  115K outer membrane protein	1387 66	609	Length	Length  283  Locus	410 Name	7.8e-40 <u>Acc#</u>	
10345327t2103  Protein name  115K outer membrane	1387 66	609	Length	Length  283  Locus	410 Name	7.8e-40 <u>Acc#</u>	
Protein name  115K outer membrane protein	1387 66	609	Length 760 2	Length  283  Locus  pir:JC6	410 3 Name 5027	7.8e-40 Acc# JC6027	
Protein name  115K outer membrane protein	1387 66	ecursor:	Length	Length  283  Locus	410 Name	7.8e-40 <u>Acc#</u>	
Protein name  115K outer membrane protein  Description	protein pre	ecursor:S	Length  760 2  SusC  NT  Length	Length  283  Locus  pir:JC6	410 3 Name 5027	7.8e-40 Acc# JC6027	
Protein name  115K outer membrane protein  Description  ORF Name	protein pre	ecursor:S	Length  760 2  SusC  NT  Length	Length  283  Locus  pir:JC6	410 3 Name 5027	7.8e-40 Acc# JC6027	
Protein name  115K outer membrane protein  Description  ORF Name  10580052_t3_183	protein pre	ecursor:S	Length  760 2  SusC  NT  Length	Length  283  Locus  pir:JC6	Allo S Name SO27	7.8e-40  Acc#  JC6027  Probability	

			$\overline{\text{NT}}$	$\underline{AA}$	Score	Probability	
ORF Name	NTID	AAID	Length	<u>Length</u>	50010		
10662877_c1_202	1389	6611	321 9	966	224	1.6e-18	
Protein name				Locus	<u>Name</u>	Acc#	
putative transposa	ıse			gp:AF00	7429	AF007429	
Description							
Haemophilus paraga	allinarum I	S-like pu	tative tr	ansposas	se gene,	complete cds.	]
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
 10725942_c3_342	1390	6612		.83			
Protein name				Locus	s Name	Acc#	
Description							
NO-HIT	×						]
	·····		NTT	7.7.			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 10819681_c2_294		AAID 6613	Length		Score	Probability  8.5e-13	
			Length	Length		- 1	
10819681 <u></u> c2 <u></u> 294			Length	Length  17  Locus	170	8.5e-13	
10819681 <u></u> c2 <u></u> 294			Length	Length  17  Locus	170 Name	8.5e-13 <u>Acc#</u>	
10819681_c2_294Protein name  Description		6613	Length	Length  17  Locus	170 Name	8.5e-13 <u>Acc#</u>	]
10819681_c2_294Protein name  Description	1391	6613	Length  138  4	Length  17  Locus  sp:MTGA	170 Name	8.5e-13 <u>Acc#</u> P44890	
10819681_c2_294Protein name  Description	1391	6613	Length	Length  17  Locus	170 Name	8.5e-13 <u>Acc#</u>	
Protein name  Description  (EC 2.4.2) (MONO	DFUNCTIONAL NTID	6613	Length  138  A  NT  Length	Length Locus sp:MTGA	170 S Name A_HAEIN	8.5e-13 <u>Acc#</u> P44890	
Protein name  Description  (EC 2.4.2) (MONO	DFUNCTIONAL NTID	G613  TGASE)  AAID	Length  138  A  NT  Length	Length Locus Sp:MTG2  AA Length	170 S Name A HAEIN Score	8.5e-13 Acc# P44890 Probability	
10819681_c2_294  Protein name  Description  (EC 2.4.2) (MONO  ORF Name  11035088_c2_306	DFUNCTIONAL NTID	G613  TGASE)  AAID	Length  138  A  NT  Length	Length Locus Sp:MTG2  AA Length	170 S Name A HAEIN Score 1634 S Name	8.5e-13 Acc# P44890  Probability  6.2e-168	
10819681c2294  Protein name  Description  (EC 2.4.2) (MONO  ORF Name  11035088c2306	DFUNCTIONAL NTID	G613  TGASE)  AAID	Length  138  A  NT  Length	Length  Locus  Sp:MTG2  AA  Length  O02  Locus	170 S Name A HAEIN Score 1634 S Name	Probability  6.2e-168  Acc#  Acc#  Acc#	

ORF Name         NTID         AAID         Length         Length         Score         Probability           Protein name         Locus Name         Accff           Description           NO-HIT           ORF Name         NTID         AAID         Length         Length         Score         Probability           ILBBL313_c3_349         1394         6616         288         867         II3         0.00062           Protein name         Locus Name         Accff           transmembrane sensor         gp:AF060193         AF060193           Description         Pseudomonas aeruginosa pigACDE operon, complete sequence; hypothetical PigB (pigB) gene, complete cds.         NTID         AAID         Length         Length         Score         Probability           I2109558_c2_2272         1395         6617         I48         447         I56         5.1e-11           Protein name         Locus Name         Accff           collagen-like protein         gp:BTU67921         U67921           Description         Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.	_							
Description   No-HIT	ORF Name	NTID	AAID			Score	Probability	
Description   NO-HIT	11832332_f1_34	1393	6615	322	969			
NO-HIT   NTID   AAID   Length   Lengt	Protein name				Locus	Name	Acc#	
ORF Name  NTID  AAID  Length  Locus Name  Acc#  Length  Description  Pseudomonas aeruginosa pigACDE operon, complete sequence; hypothetical PigB (pigB) gene, complete cds.  ORF Name  NTID  AAID  Length  Length  Length  Length  Length  Locus Name  Acc#  Collagen-like protein  Description  Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.  ORF Name  NTID  AAID  Length  Length  Length  Length  Length  Length  Score  Probability  Description  Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.	Description							
ORF Name  NTID  AAID  Length								
ORF Name  NTID  AAID  Length								-
Protein name  Locus Name Acc#  transmembrane sensor  Description  Pseudomonas aeruginosa pigACDE operon, complete sequence; hypothetical PigB  (pigB) gene, complete cds.  ORF Name  NTID AAID Length Length Length Score Probability  1210.9558c2_2.72	ORF Name	NTID	AAID			Score	Probability	
Description  Pseudomonas aeruginosa pigACDE operon, complete sequence; hypothetical PigB (pigB) gene, complete cds.  ORF Name  NTID  AAID  Length  Length  Length  Score  Probability  12109558_c2_272	11881313c3349	. 1394	6616	288	367	113	0.00062	
Description  Pseudomonas aeruginosa pigACDE operon, complete sequence; hypothetical PigB (pigB) gene, complete cds.  NTI AA Score Probability  Description  Protein name  Locus Name Acc#  Collagen-like protein  Description  Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.  NTID AAID Length Length Score Probability  Description  Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.	Protein name				Locus	Name	Acc#	
Pseudomonas aeruginosa pigACDE operon, complete sequence; hypothetical PigB (pigB) gene, complete cds.    NT	transmembrane senso	or			gp:AF06	0193	AF060193	
(pigB) gene, complete cds.    NTID   AAID   Length   Leng	Description							
ORF Name  NTID  AAID  Length  Length  Score  Probability  12109558_c2_2721395  Frotein name  Locus Name  Acc#  Collagen-like protein  Description  Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.  NTID  AAID  NT  AA  Score  Probability  Probability  AAID  Length  Length  Length  Score  Probability  Probability			E operon,	complet	ce sequer	ice; hypot	thetical PigB	
ORF Name  NTID  AAID  Length  Length  Score  Probability  12109558_c2_2721395  Frotein name  Locus Name  Acc#  Collagen-like protein  Description  Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.  NTID  AAID  NT  AA  Score  Probability  Probability  AAID  Length  Length  Length  Score  Probability  Probability								
Protein name  Locus Name  Acc#  collagen-like protein  Description  Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.  NTID  AAID  Length  Length  Score  Probability	ORF Name	NTID	AAID			Score	Probability	
Collagen-like protein  Description  Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.  ORF Name  NTID  AAID  Length  Length  Score  Probability	12109558 <u></u> c2272	. 1395	6617	148	147	156	5.1e-11	
Description  Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.  ORF Name  NTID  AAID  Length  Length  Probability	Protein name				Locus	Name	Acc#	
Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.  NT AA  ORF Name  NTID AAID Length Length  Probability	collagen-like prote	ein			gp:BTU6	7921	U67921	
genes, complete sequence.  NT AA  ORF Name  NTID AAID Length Length  Probability	Description							
ORF Name NTID AAID Length Length Probability	_	_	d pTX14-	L, MOB, I	REP, and	collage	n-likeprotein	
ORF Name NTID AAID Length Length Score Probability	gener, compress seq							
ORF Name NTID AAID Length Length				NT	AA	Cacro	Drobability.	
1271010 62 297 1396 16618 196 1291	ORF Name	NTID	AAID	Length	Length	Score	Probability	
1370 0010 271	1271010 <u></u> .c3 <u></u> 387	. 1396	6618	96	291			
Protein name Locus Name Acc#	Protein name				Locus	Name	Acc#	
	Description							
<u>Description</u>	NO-HIT							

			$\underline{ ext{NT}}$	<u>AA</u>	Score	Probability	
ORF Name	NTID	AAID	Length	<u>Length</u>	<u> </u>		
13071943_c1_216	1397	6619	466	.401	397	7.5e-37	
Protein name				Locus	3 Name	Acc#	
conserved hypothe	tical protein	n		pir:H72	2331	H72331	
Description							
ODE Name	ATOM TO IN	* * TD	NT	<u>AA</u>	Score	Probability	
ORF Name	NTID	AAID	Length	Length			
13.750800 <u></u> c120.7	1398	6620	78 2	:37			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT	·				-	-	
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
13806517 <u></u> £2 <u></u> 85	1399	6621	401 1	206	169	1.0e-09	
Protein name	j			Locus	Name	Acc#	
transposase		· ¥ ·	*	gp:AF03	88866	AF03886	5
Description	The state of the s	10					
Bacteroides fragi protein BmpH (bmpH			_	sase (bi	pH) andm	obilization	
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
14469691 <u></u> c1213	1400	6622	193 5	82	199	7.2e-16	
Protein name				Locus	Name	Acc#	
RNA polymerase sig	gma factor Si	igZ-like	protein	gp:AF13	7263	AF137263	3
Description		*		· · · · · · · · · · · · · · · · · · ·	1	(i)	
Bacteroides theta gene cluster, and complete cds.			<del>-</del>		~		,

ORF Name	NTID A	<u>NT</u> AAID Length	AA Length Score	Probability
14589067_f3_150	1401 662	93	282	
Protein name			Locus Name	Acc#
Description				
NO-HIT				
		NT	AA	
ORF Name	NTID I	AID Length	— Score	Probability
14663552f277	1402 662	296	891 116	0.00021
Protein name			Locus Name	Acc#
			sp:LCRF_YERPE	P28808
Description			-	
	COURTNE LODE			
THERMOREGULATORY PI	KOIEIN ICRE			
THERMOREGULATORY PI	COLETN LCKE			
ORF Name		<u>NT</u> AID Length	AA Length Score	Probability
	NTID I	AID Length	Score	Probability  0.029
ORF Name	NTID I	AID Length	Length Score	*
ORF Name 14723156t2132	<u>NTID</u> 2	AID Length	Length Score  1372 83	0.029
ORF Name  14723156£2132  Protein name	<u>NTID</u> 2	AID Length	Length Score  Length Score  B372 83  Locus Name	0.029 <u>Acc#</u>
ORF Name  14723156t2132  Protein name  hypothetical protein	<u>NTID</u> 2	AID Length	Length Score  Length Score  B372 83  Locus Name	0.029 <u>Acc#</u>
ORF Name  14723156t2132  Protein name  hypothetical protein	NTID 4	AID Length	Length Score  Length Score  B372 83  Locus Name  pir:H70478	0.029 <u>Acc#</u>
ORF Name  14723156f2132  Protein name  hypothetical protein  Description	NTID 4	AID Length  123  NT Length Length	Length Score  Length Score  B372 83  Locus Name  pir:H70478	0.029 <u>Acc#</u> H70478
ORF Name  14723156f2132  Protein name  hypothetical protei  Description  ORF Name	NTID 4	AID Length  123  NT Length Length	Length Score  Length Score  B372 83  Locus Name  pir:H70478  AA Score  Length Score	0.029 <u>Acc#</u> H70478
ORF Name  14723156f2132  Protein name  hypothetical protei  Description  ORF Name  14875302c1267	NTID 4	AID Length  123  NT Length Length	Length Score  Length Score  372 83  Locus Name  pir:H70478  AA Length Score  489	Acc# H70478  Probability
ORF Name  14723156f2132  Protein name  hypothetical protei  Description  ORF Name  14875302c1267  Protein name	NTID 4	AID Length  123  NT Length Length	Length Score  Length Score  372 83  Locus Name  pir:H70478  AA Length Score  489	Acc# H70478  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
15659758_£1_51	1405	627	63 1	.92		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
15671890c3394	1406	628	428 1	287	191	3.9e-12
Protein name				Locus	Name	Acc#
transposase				gp:AF03	8866	AF038866
Description		•		•		
Bacteroides fragi protein BmpH (bmpH	_		_	sase (bi	pH) andr	nobilization
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 15736057±129		AAID	<u>Length</u>		Score	Probability 2.2e-211
			<u>Length</u>	Length	-	
15736057 <u></u> £1 <u></u> 29			<u>Length</u>	Length	2044 Name	2.2e-211
15736057 <u></u> £1 <u></u> 29			<u>Length</u>	Length 572 Locus	2044 Name	2.2e-211 <u>Acc#</u>
15736057 <u></u> ±1 <u></u> 29 Protein name		629	Length 523	Length  572  Locus  sp:TRA2	2044 Name	2.2e-211 <u>Acc#</u>
Protein name  Description		629	Length 523	Length  572  Locus  sp:TRA2	2044 Name	2.2e-211 <u>Acc#</u>
Protein name  Description  TRANSPOSASE FOR I	NSERTION SEQU	JENCE ELE	Length  523  EMENT IS2  NT  Length	Length  572  Locus  Sp:TRA2  1-LIKE  AA	Name BACFR	2.2e-211 <u>Acc#</u> Q45119
Protein name  Description  TRANSPOSASE FOR I	NSERTION SEQU	JENCE ELE	Length  523  EMENT IS2  NT  Length	Length  572  Locus  Sp:TRA2  1-LIKE  AA  Length  437	Name BACFR Score	2.2e-211 <u>Acc#</u> Q45119  Probability
Protein name  Description  TRANSPOSASE FOR I	NSERTION SEQU	JENCE ELE	Length  523  EMENT IS2  NT  Length	Length  572  Locus  Sp:TRA2  1-LIKE  AA  Length  437	2044 Name BACFR  Score 468 Name	2.2e-211  Acc# Q45119  Probability  2.2e-44
Protein name  Description  TRANSPOSASE FOR I	NSERTION SEQU	JENCE ELE	Length  523  EMENT IS2  NT  Length	Length  572  Locus  Sp:TRA2  1-LIKE  AA  Length  437  Locus	2044 Name BACFR  Score 468 Name	2.2e-211  Acc# Q45119  Probability  2.2e-44  Acc#

ORF Name NTID AAID Length Length $\frac{AA}{AB}$ Score Probabili	ty
16491593_c2_279	
Protein name Locus Name Acc#	
	•
<u>Description</u>	
NO-HIT	
ORF Name NTID AAID Length Length Score Probabili	ty
19531438f149	
Protein name Locus Name Acc#	
Description	
NO-HIT	
NT AA C. Duckelelele	
ORF Name NTID AAID Length Length Score Probabili	ty
1993.763.7 <u>t2111</u>	
Protein name Locus Name Acc#	
Description	
NO-HIT	
NT AA Garage Bushahili	
ORF Name NTID AAID Length Length Score Probabili	<u>ty</u>
2145012t3156 1412 6634 67 204 95 0.00024	
Protein name Locus Name Acc#	-
hypothetical 26.8K protein pir:JC2322 JC232	:2
Description	
NT AA Grana Brokahili	
ORF Name NTID AAID Length Length Score Probabili	<u>ty</u>
216.83.280 <u></u> c33.50	
Protein name Locus Name Acc#	
Description	
NO-HIT	

ORF Name	NTID AAID	NT AA Length Length Score Probability	
22459687_c3_347	1414   6636	187  564	
Protein name		Locus Name Acc#	
Description			
NO-HIT			]
ORF Name	NTID AAID	NT AA Length Length Score Probability	
22691552a1262	1415 6637	434 1305 140 2.7e-06	
Protein name	-	Locus Name Acc#	
immunoreactive 53	kD antigen PG123	gp:AF144641 AF144641	
Description	*		
Porphyromonas ging complete cds.	givalis strain W50 :	Immunoreactive 53 kD antigenPG123 gene,	
ORF Name	NTID AAID	NT AA Score Probability	_
22692067 <u></u> c1233	1416 6638	324 975 445 6.1e-42	
Protein name		Locus Name Acc#	
		sp:HTPX_STRGC 030795	
Description			
PUTATIVE HEAT SHOO	CK PROTEIN HTPX		]
		NT AA Company	
ORF Name	NTID AAID	Length Length Score Probability	
22836061 <u></u> £2 <u></u> 82	1417 6639	133 402 97 0.00018	
Protein name		Locus Name Acc#	
МррВ		gp:BFU25716 U25716	
Description			

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
22933438_c3_397	1418	6640	322	969		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
		* 2 TD	NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length	×	
22933438f121	1419	6641	242	729	Mana	7 4
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
					-	
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
ORF Name 23495700_f2_138	NTID 1420	AAID	Length		Score	Probability
			Length	Length	Score Name	Probability  Acc#
23495700 <u>f2</u> 138			Length	Length	× .	*
23495700f2138 Protein name			Length	Length	× .	*
23495700_f2_138 Protein name  Description		6642	Length 67 2	Length 204 Locus	Name	Acc#
23495700_f2_138 Protein name  Description			Length	Length 204 Locus	× .	*
23495700f2138  Protein name  Description  NO-HIT  ORF Name  23531265f3149	1420	6642	Length 67  NT Length	Length Locus  AA Length	Name Score	Acc#
23495700f2138  Protein name  Description  NO-HIT  ORF Name	NTID	AAID	Length 67  NT Length	Length Locus  AA Length	Name	Acc#
23495700f2138  Protein name  Description  NO-HIT  ORF Name  23531265f3149	NTID	AAID	Length 67  NT Length	Length Locus  AA Length	Name Score	Acc# Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
23648392 ±2_86	1422	6644		296	141	2.1e-06
			<u> </u>			
Protein name				Locus	Name	Acc#
immunoreactive 53	kD antigen	PG123		gp:AF14	14641	AF144641
Description						
Porphyromonas gin	givalis stra	ain W50 ir	mmunoreac	ctive 53	kD antig	genPG123 gene,
complete cds.		*				
	*		NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length	BCOLE	riobability
23679510_c3_345	1423	6645	197	594	345	2.4e-31
Protein name				Locus	s Name	Acc#
putative acetyltra	ansferase			gp:SCF	L	AL117322
Description						
Streptomyces coel	- aclas acam:	2.3. 114				
Screptomyces coer	TCOTOL COSM.	la Fi.			7	
Screptomyces coer	ICOTOT COSIII.	1.0 F1.	NT	AA		
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
	NTID		<u>Length</u>		Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	Length	Score S Name	Probability  Acc#
ORF Name  24026502_f1_27  Protein name	NTID	AAID	<u>Length</u>	Length		
ORF Name  24026502_f1_27  Protein name  Description	NTID	AAID	<u>Length</u>	Length		
ORF Name  24026502_f1_27  Protein name	NTID	AAID	<u>Length</u>	Length		
ORF Name  24026502_f1_27  Protein name  Description	NTID	AAID	<u>Length</u>	Length		
ORF Name  24026502_f1_27  Protein name  Description  NO-HIT	NTID 1424 NTID	<u>AAID</u>  6646	Length  88  NT Length	Length Locus	s Name	Acc#
ORF Name  24026502_f1_27  Protein name  Description  NO-HIT  ORF Name	NTID 1424 NTID	AAID 6646 AAID	Length  88  NT Length	Length Locus  AA Length	Score	Acc# Probability
ORF Name  24026502_f1_27  Protein name  Description  NO-HIT  ORF Name  24259637c3_336	NTID 1424 NTID	AAID 6646 AAID	Length  88  NT Length	Length Locus  AA Length	Score 253 S Name	Acc# Probability 3.1e-37
ORF Name  24026502_f1_27  Protein name  Description  NO-HIT  ORF Name  24259637_c3_336  Protein name	NTID 1424 NTID	AAID 6646 AAID	Length  88  NT Length	Length Locus  AA Length 2073 Locus	Score 253 S Name	Acc#  Probability  3.1e-37  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24332035_c1_197  Protein name	1426	6648	83	Locus	s Name	Acc#	
<u>Description</u>				<u>посак</u>	<u> Name</u>	120#	
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24407551 <u></u> c2 <u></u> 293	1427	6649	231	96	427	5.0e-40	]
Protein name				Locus	s Name	Acc#	
immunogenic 23 kDa	lipoprotei	in PG3		gp:AF14	15799	AF145799	
Description							
Porphyromonas ging	valis stra	ain W50 i	mmunogen	LC 23 kDa	a lipopro	oteinPG3 gene,	
complete cds.							
complete cas.			NT	AA			0
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	0
	NTID	<u>AAID</u>	Length		Score	Probability	
ORF Name			Length	Length	Score S Name	Probability  Acc#	
ORF Name  24415757c3348			Length	Length			
ORF Name  24415757_c3_348  Protein name			Length	Length			
ORF Name  24415757_c3_348  Protein name  Description			Length	Length			
ORF Name  24415757_c3_348  Protein name  Description  NO-HIT	NTID	6650	Length  424  NT Length	Length Locus	s Name	Acc#	
ORF Name  24415757c3348  Protein name  Description  NO-HIT  ORF Name	NTID	AAID	Length  424  NT Length	Length Locus  AA Length	Score	Acc# Probability	]
ORF Name  24415757c3348  Protein name  Description  NO-HIT  ORF Name  24641061£296	NTID	AAID	Length  424  NT Length	Length Locus  AA Length	Score 147 S Name	Acc# Probability 9.5e-08	]

ORF Name	NTID AAID	NT AA Score Probability Length Length
24642137_£1_62	1430   6652	401 1206 176 2.9e-11
Protein name		Locus Name Acc#
putative outer me	embrane porin	gp:AF030977
Description		
outer membrane por	cin (ompA), unknown p	ase (gltX) gene, partial cds;putative protein, vibriobactinreceptor precursor completecds; and VibF (vibF) gene,
ORF Name	NTID AAID	NT AA Score Probability Length Length
24642212 <u>f2</u> 110	1431   6653	
Protein name		Locus Name Acc#
FIOCEIII IIdiiic		sp:YBFH BACSU 031448
Description		<u>-</u>
14	8 KD PROTEIN IN GLPT	PURT INTERGENIC REGION
		NIT AA
ORF Name	NTID AAID	
ORF Name 24647826t128		NT AA Score Probability
		NT AA Score Probability
24647826 <u></u> £128		NT AA Score Probability Length Length 234
24647826t128 Protein name		NT AA Score Probability Length Length 234
Protein name  Description		NT AA Score Probability    77   234
Protein name  Description		NT AA Score Probability Length Length  77 234  Locus Name Acc#
Protein name  Description  NO-HIT	1432   6654   6654   MTID AAID	NT AA Score Probability  Length Length Score Probability  Locus Name Acc#
Protein name  Description  NO-HIT  ORF Name	1432   6654   6654   MTID AAID	NT AA Score Probability  77 234  Locus Name Acc#  NT AA Score Probability  NT AA Score Probability
Protein name  Description  NO-HIT  ORF Name  24725380_c1_245	NTID AAID [1433] [6655]	NT AA Score Probability    77   234
Protein name  Description  NO-HIT  ORF Name  24725380c1245  Protein name  mobilization production	NTID AAID  1433 6655  tein A	NT AA Score Probability    77   234

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24726592_c3_355	1434	6656		159	339	1.0e-30	
Protein name			<u> </u>	Locus	Name	Acc#	
				sp:MTG/	ECOLI	P46022	
Description						ý.	
(EC 2.4.2) (MON	OFUNCTIONAL	TGASE)					
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24803426_c1_210	1435	6657	204	515	120	1.9e-06	
Protein name				Locus	Name	Acc#	
hypothetical prote	ein MTH847			pir:A69	9213	A69213	
Description							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24847551 <u>c2</u> 302	1436	6658	92	279		,	
Protein name				Locus	s Name	Acc#	
Description							
NO-HIT						:	
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
		<u>AAID</u>	Length		Score	Probability	
ORF Name			Length	Length		1	
ORF Name 2538277c2282	1437		Length	Length	120 S Name	1.7e-07	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
25511052_c1_232	1438	6660	206	21	401	2.8e-37
Protein name				Locus	Name	Acc#
LemA				gp:LMU6	6186	U66186
Description						
Listeria monocytog partial cds.	genes LemA	(lemA) gei	ne, compl	ete cds,	and Len	nB(lemB) gene,
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
25527053_f1_22	1439	6661	436	[311		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						-
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
		<u>AAID</u>	Length		Score	Probability
ORF Name			Length	Length	Score S Name	Probability  Acc#
ORF Name 26354518£295			Length	Length	*	
ORF Name  26.3.5.4.5.1.8			Length	Length	*	
ORF Name  26354518f295  Protein name  Description			Length	Length	3 Name	Acc#
ORF Name  26354518f295  Protein name  Description			Length	Length  2222  Locus	*	Acc# Probability
ORF Name  26354518f295  Protein name  Description  NO-HIT		6662	Length  73  NT Length	Length  Locus	3 Name	Acc#
ORF Name  26354518t295  Protein name  Description  NO-HIT  ORF Name		AAID	Length  73  NT Length	Length Locus  AA Length	Score	Acc# Probability
ORF Name  26354518t295  Protein name  Description  NO-HIT  ORF Name  2767137c1266	NTID	<u>AAID</u>	Length  73  NT Length	Length Locus  AA Length	Score 128 Name	Acc# Probability  0.00031

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
	1442	6664	<del></del>	:64		
2792942_£2_70	1442	0004				- "
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
29713458f283	1443	6665	67 2	204		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA		D
ORF Name	NTID	AAID	Length	Length	Score	Probability
31923438 <u></u> c2 <u></u> 291	1444	6666	61.	L86		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NTID	7.7.		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
32213312 <u></u> t3 <u></u> 163	1445	6667	171	516	237	6.8e-20
Protein name				Locus	Name	Acc#
putative ECF sigma	factor Rp	ooE1		gp:AF04	9107	AF049107
Description						* <u>-</u>
Myxococcus xanthus dehydrogenase (aldA)	, putativ	re ECF sig	ma facto	r RpoE1(r	poEl), a	and response

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
33214538_£3_155	1446	6668	744 2	235		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA		
ORF Name	NTID	AAID	Length	<u>AA</u> Length	Score	Probability
33486716 <u></u> c2 <u></u> 276	[1447	6669	497 1	494	793	8.2e-79
Protein name				Locus	Name	Acc#
				sp:HEMM	N_AQUAE	067886
Description						
OXYGEN-INDEPENDEN	r coproporpi	HYRINOGEN	II			, 1
			NTT	77		
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 33632692_c1_229	-	<u>AAID</u>	Length		Score	Probability
•	-		Length	Length	Score S Name	Probability  Acc#
33632692 <u>c1229</u>	-		Length	Length		*
33632692_c1_229 Protein name	-		Length	Length		*
Protein name  Description	-		Length 71 2	Length 216 Locus		*
Protein name  Description	-		Length	Length		*
Protein name  Description  NO-HIT	1448 NTID	6670	Length  71  2  NT  Length	Length Locus	s Name	Acc#
Protein name  Description  NO-HIT  ORF Name	1448 NTID	AAID	Length  71  2  NT  Length	Length Locus  AA Length	Score	Acc# Probability
33632692_c1_229  Protein name  Description  NO-HIT  ORF Name  34181502_f2_122	NTID [1449	AAID	Length  71  NT  Length  426  1	Length Locus  AA Length	Score 134 Name	Acc# Probability 6.9e-06

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
35302 f1_52	1450	6672	203 6	12			
			L L	T.OCUS	. Name	Acc#	
Protein name				посия	, Ivallic	110011	
Description							
NO-HIT							
			NT	AA	Caana	Drobabilita	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
3942130 <u>c1264</u>	1451	6673	388 1	.167	86	0.0055	]
Protein name				Locus	s Name	Acc#	•
integrase				gp:HIV	J69223	U69223	
Description				_			
HIV-1 strain CMR27	3 from Cam	eroon int	egrase (p	ool) gene	e, partia	alcds.	
			NT	AA			_
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability	
3.954762f3187	1452	6674	291	376	11.1	6.4e-06	]
Protein name				Locu	s Name	Acc#	
transcription regu	lator homo	log:hypot	netical	pir:PC	4110	PC4110	
137 protein							
Description							
	:		NTT	AA			_
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4114055_c3_337	1453	6675	415	1248	288	5.0e-24	]
Protein name				Locu	s Name	Acc#	
hypothetical prote	ein			gp:AF1	49851	AF149851	
Description					-		
Pseudomonas sp. Ko	d himothet:	ical profe	ins. met	hallothi	onein-li	keprotein,	
MoeR-like protein	nutative r	oroteins.	hypothet	icalprot	ein, put	ative	
lavidamaduataga an							
putative receptor	d putative	AMP ligas	e (entE)	genes, c	omplete	cds; and	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4720187_±2_99	1454	6676	283	I.OCUS	954 S Name	7.1e-96 Acc#	
<u>Protein name</u>				sp:IST		Q45120	
Description  INSERTION SEQUE	NCE IS21-LIKE	PUTATIVE	ATP-BINI	OING PROT	EIN	*	1
INDURITION DEGEL							]
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4822751_f2_101	1455	6677	594	785	370	1.8e-48	
Protein name					s Name	Acc#	
oxaloacetate de (oadA) homolog	carboxylase,	subunit a.	lpha	pir:C6	9406	C69406	
Description							
			NT	AA	Score	Probability	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 4864702_t2_124		<u>AAID</u>	Length		Score	Probability	
		-	Length	Length	Score S Name	Probability  Acc#	10
4864702 <u></u> £2 <u></u> 124		-	Length	Length			÷
4864702_f2_124 Protein name		-	Length	Length			
Protein name  Description		-	Length	Length	s Name	Acc#	]
Protein name  Description		-	Length	Length  Locus			]
Protein name  Description  NO-HIT		6678	Length  152  NT Length	Length Locus	s Name	Acc#	]
Protein name  Description  NO-HIT  ORF Name		6678 AAID	Length  152  NT Length	Length Locus  AA Length	Score	Acc# Probability	
Protein name  Description  NO-HIT  ORF Name  4879635f161		6678 AAID	Length  152  NT Length	Length Locus  AA Length	Score 214 s Name	Acc# Probability 4.6e-14	
Protein name  Description  NO-HIT  ORF Name  48.796.35_f1_61  Protein name		6678 AAID	Length  152  NT Length	Length Locus  AA Length  Locus  Locus  Locus	Score 214 s Name	Acc#  Probability  4.6e-14  Acc#	, 

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
5177157_f2_88  Protein name	1458	6680	182	Locus	Name	Acc#
Description NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
5285692 <u></u> c2 <u></u> 281	1459	6681	86	261		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NID	7.7		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
5321932 <u></u> ±1 <u></u> 53	1460	6682	234	705		
Protein name				Locus	Name	Acc#
Description						
NO-HIT	·	0				
			NT	AA		
ORF Name	$\underline{\mathtt{NTID}}$	AAID	Length	Length	Score	Probability
5.6.4.0.3.7c33.4.6	1461	6683	282	349	229	4.8e-19
Protein name				Locus	Name	Acc#
hypothetical protein	1	· · · · · ·	4	pir:B72	308	B72308
Description						
			NT	AA		Dec - 1 - 1 - 2 7 - 2 to
ORF Name	NTID	$\underline{\mathtt{AAID}}$	<u>Length</u>	Length	Score	Probability
5970252 <u></u> c2 <u></u> 316	1462	6684	61	186		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
				*			
6025010_c3_333	1463	6685	74	225			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT				*			
							_
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
6.046.90.7£13.5	1464	6686	636	1911			
Protein name	*			Locus	Name	Acc#	
Description							
NO-HIT			:				
		*					
			NTT	Z\ Z\			
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
ORF Name 6.7.7.5.4.3.8	NTID	<u>AAID</u>	Length		Score	Probability	
			Length	Length	Score Name	Probability  Acc#	
67.75438 <u></u> c3 <u></u> 401			Length	Length			
6.7.7.5.43.8			Length	Length			
Protein name  Description			Length	Length  375  Locus			
Protein name  Description			Length	Length			
Protein name  Description  NO-HIT	1465	6687	Length  124  NT  Length	Length  375  Locus	s Name	Acc#	]
Protein name  Description  NO-HIT  ORF Name	NTID	AAID	Length  124  NT  Length	Length  Locus  AA  Length	Name Score	Acc# Probability	
Description  NO-HIT  ORF Name  785322_t1_64	NTID	AAID	Length  124  NT  Length	Length  Locus  AA  Length	Score 220 Name	Acc# Probability 2.7e-20	
Description  NO-HIT  ORF Name  785322_t1_64	NTID	AAID	Length  124  NT  Length	Length  Locus  AA  Length  1752  Locus	Score 220 Name	Acc#  Probability  2.7e-20  Acc#	]

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
15808290_c1_33	1467	6689	61 1	.86	99	2.8e-05
Protein name				Locus	Name	Acc#
glycine-rich prote	ein (clone w	710-1)		pir:S14	982	S14982
Description						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
19659503_c1_32	1468	6690	383	152		
Protein name		1	*	Locus	Name	Acc#
Description						
NO-HIT			· · · · · · · · · · · · · · · · · · ·			
			NT	AA		
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 26367141_c1_30		<u>AAID</u>	Length		Score	Probability  3.5e-05
			Length	Length		
26367141_c1_30	1469		Length	Length	130 Name	3.5e-05
26367141_c1_30 Protein name	1469		Length	Length 379 Locus	130 Name	3.5e-05 <u>Acc#</u>
Protein name  membrane glycopro	1469 tein	6691	Length  292	Length  B79  Locus  gp:D887	130 Name	3.5e-05 Acc# D88733
Protein name  membrane glycopro  Description	1469 tein	6691	Length  292  [Signature of the content of the conte	Length  Locus  gp:D887	Name  //33	3.5e-05 <u>Acc#</u> D88733
Protein name  membrane glycopro  Description	1469 tein	6691	Length  292	Length  B79  Locus  gp:D887	130 Name	3.5e-05 Acc# D88733
Protein name  Membrane glycopro  Description  Equine herpesviru	tein s 1 DNA for	membrane	Length  292  glycopro  NT  Length	Length  Locus  Jop: D887  Dtein, co	Name  //33	3.5e-05 <u>Acc#</u> D88733
Protein name  Protein name  membrane glycopro  Description  Equine herpesviru  ORF Name	tein s 1 DNA for	membrane	Length  292  glycopro  NT  Length	Length Locus Locus AA Length Length	Name  //33	3.5e-05 <u>Acc#</u> D88733
Protein name  Protein name  membrane glycopro  Description  Equine herpesviru  ORF Name	tein s 1 DNA for	membrane	Length  292  glycopro  NT  Length	Length Locus Locus AA Length Length	Name  /33  mplete o	3.5e-05  Acc#  D88733  Eds.  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
34666302_c3_43		<del></del>		488	213	1.3e-20
Protein name			L	Locus	Name	Acc#
immunoreactive 5	3 kD antigen P	G123		gp:AF14	4641	AF144641
Description						
Porphyromonas gi complete cds.	ngivalis strai	n W50 im	munoreac	tive 53	kD antig	genPG123 gene,
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
807033_c1_29	1472 6	694	112 3	39		
Protein name				Locus	Name	Acc#
Description						
NYA TYPH						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
		<u>AAID</u>	Length	<del></del>	<u>Score</u> 452	Probability
ORF Name			Length	Length	*	
ORF Name 11063391_c2_40  Protein name			Length	Length	452 Name	1.1e-42
ORF Name  11063391_c2_40  Protein name  Description	 [1473		Length	Length  035  Locus	452 Name	1.1e-42 Acc#
ORF Name 11063391_c2_40  Protein name	 [1473		Length	Length  035  Locus	452 Name	1.1e-42 Acc#
ORF Name  11063391_c2_40  Protein name  Description	 [1473		Length	Length  035  Locus	452 Name	1.1e-42 Acc#
ORF Name  11.06.33.91c240  Protein name  Description  HYPOTHETICAL PRO	TEIN HI1523	AAID	Length  344  NT Length	Length  035  Locus  sp:YF23	452 Name HAEIN	1.1e-42 Acc# P44243
ORF Name  11.06.33.91c240  Protein name  Description  HYPOTHETICAL PRO	TEIN HI1523	AAID	Length  344 1  NT Length	Length  035  Locus  Sp:YF23  AA  Length	452 Name HAEIN	1.1e-42 Acc# P44243
ORF Name  11.06.33.91_c2_40  Protein name  Description  HYPOTHETICAL PRO  ORF Name  14.348.958_f2_14	TEIN HI1523	AAID	Length  344 1  NT Length	Length  035  Locus  Sp:YF23  AA  Length	A52 Name HAEIN Score	Acc# P44243  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
14650012_f3_26	1475	6697	145 4	35	73	0.022	
Protein name				Locus	Name	Acc#	
glucosidase II bet	a-subunit			gp:AF06	6061	AF066061	
Description							
Mus musculus gluco products, partial o		oeta-subu	nit gene,	alterna	tivelysp	oliced	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
15835261_c1_28	1476	6698	116	51			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							1
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability	-
ORF Name 23472832_c1_31		AAID 6699	Length		Score	Probability 6.8e-43	
ORF Name  23472832_c1_31  Protein name			Length	Length		3	
23472832 <u></u> c1 <u></u> 31			Length	Length	454 s Name	6.8e-43	
23472832 <u></u> c1 <u></u> 31			Length	Length Locus	454 s Name	6.8e-43 <u>Acc#</u>	
23472832_c1_31 Protein name	1477	6699	Length	Length Locus	454 s Name	6.8e-43 <u>Acc#</u>	
Protein name  Description	(BILE SALT	6699 HYDROLAS	Length  366	Length Locus sp:CBH	454 s Name	6.8e-43 <u>Acc#</u>	]
Protein name  Description  HYDROLASE) (CBAH)  ORF Name	(BILE SALT	HYDROLAS AAID	Length  366  E  NT  Length	Length  Locus  Sp:CBH  AA  Length	454  Name CLOPE	6.8e-43 <u>Acc#</u> P54965	
Protein name  Description  HYDROLASE) (CBAH)  ORF Name	(BILE SALT	6699 HYDROLAS	Length  366  NT Length	Length  Locus  Sp:CBH  AA  Length	A54 S Name CLOPE Score	6.8e-43 Acc# P54965 Probability	
Protein name  Description  HYDROLASE) (CBAH)  ORF Name	(BILE SALT	HYDROLAS AAID	Length  366  E  NT  Length	Length  Locus  Sp:CBH  AA  Length	454  Name CLOPE	6.8e-43 <u>Acc#</u> P54965	
Protein name  Description  HYDROLASE) (CBAH)  ORF Name	(BILE SALT	HYDROLAS AAID	Length  366  E  NT  Length	Length  Locus  Sp:CBH  AA  Length	A54 S Name CLOPE Score	6.8e-43 Acc# P54965 Probability	

ORF Name NTID AAID Length Length Score Probability	
24798401_f1_2	
Description	
NO-HIT	
NT AA Score Probability	_
ORF Name NTID AAID Length Length	
3.03.5.5.3.0.5£324	j
Protein name Locus Name Acc#	
sp:SOJ_BACSU P37522	
Description	
SOJ PROTEIN	
ORF Name NTID AAID Length Length $\frac{NT}{N}$ Score Probability	
3388283.7£325	,
	ļ
Protein name Locus Name Acc#  hypothetical protein F20D10.230 pir:T05638 T05638	
Description	
ORF Name NTID AAID Length Length Score Probability	
4407575_£2_121482 6704 522 1569 115 2.8e-10	
Protein name Locus Name Acc#	
endo-xylanase homolog PCZA361.14 pir:T17480 T17480	
Description	
ORF Name NTID AAID Length Length Score Probability	
586.916.7£17	
Protein name Locus Name Acc#	
Description	
NO-HIT	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
6838437_f1_1	1484	6706	422	.269	100	0.0024	
Protein name				Locus	Name	Acc#	
outer membrane pro	otein			gp:BNRC	MPB	L77614	
Description				-			
Bacteroides theta cds.	iotaomicron	outer men	mbrane pr	otein (s	usD) gei	ne,complete	
			NT	AA	Ccoro	Drobability	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
10547256_f2_6	1485	6707	81 2	46			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
11126552 <u>c214</u>	1486	6708	152 4	59	95	7.5e-05	
Protein name	1486	6708	152	*	95 Name	7.5e-05 Acc#	
		-	152 4	*	Name		
Protein name		-	152 [4	Locus	Name	Acc#	
Protein name hypothetical protein		-		Locus	Name	Acc#	,
Protein name hypothetical protein		-	NT Length	Locus	Name	Acc#	
Protein name  hypothetical protein  Description	ein aq_1018		NT Length	Locus pir:H70	Name	<u>Acc#</u> H70387	7
Protein name  hypothetical protein  Description  ORF Name	ein aq_1018	AAID	NT Length	Locus pir:H70  AA Length	Name	<u>Acc#</u> H70387	7
Protein name  hypothetical protein  Description  ORF Name  2117841t11	ein aq_1018	AAID	NT Length	Locus pir:H70  AA Length	Score	Acc# H70387 Probability	

ORF Name	NTID AAID	NT <u>AA</u> Score Length Length	Probability
26682828_c3_19	1488 6710	136 411 93	0.023
Protein name		Locus Name	Acc#
surface exclusion	protein sepl precur	sor pir:S72375	S72375
Description			
		NT AA Score	Probability
ORF Name	NTID AAID	<u>Length</u> <u>Length</u>	
31417187 <u>c1</u> 10	1489 6711	261 786	7
Protein name		<u>Locus Name</u>	Acc#
Description		-	-
NO-HIT			*
ORF Name	NTID AAID	NT AA Score	Probability
6.03.568.7c218	-	]	8.3e-141
Protein name		Locus Name	Acc#
fructanase		pir:A36915	A36915
Description			
		NT AA Score	Probability
ORF Name	NTID AAID	Length Length	
85912 <u></u> c1 <u></u> 11	1491 6713	790 2373 2229	5.5e-231
Protein name		Locus Name	Acc#
		gp:BNRSCRL	M83774
Description	lis levanase (scrl)	gene, complete cds.	
bacterordes fragi	TID ICVARIABLE (SEEL)		
ORF Name	NTID AAID	NT AA Score	Probability
14333277 <u>c1</u> 10	1492 6714	127 384	
Protein name	v.	Locus Name	Acc#
Description			
NO-HIT			

			NT	AA Taranah	Score	Probability
ORF Name	NTID	AAID	Length	Length		
15735882_£3_5	1493	6715	400	.203	696	1.5e-68
Protein name				Locus	s Name	Acc#
renin-binding prote slr1975:protein slr1		d protein	:protein	pir:S75	5649	S75649
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
26819566 <u></u> £3 <u></u> 4	1494	6716	73	222		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT						
		,	NT	AA	Saoro	Probability
ORF Name	NTID	AAID	Length	Length	Score	FIODADITICY
31381 <u>f3</u> 6	1495	6717	417	L254	280	8.7e-23
Protein name				Locus	s Name	Acc#
hexuronate transpor	rter homol	og yjmG		pir:A6	9853	A69853
Description					/	
				7.7		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
3956707 <u>_</u> ±1 <u>_</u> 1	1496	6718	149	150	126	1.4e-07
Protein name				Locus	s Name	Acc#
N-acetylneuraminate	e lyase			gp:CPN	ANA	Y12876
Description				Ĭ.		
C.perfringens gene reading frames.	encoding	N-acetyln	euramina	te lyase	and two	partial open

			NT	AA	Score	Probability	
ORF Name	NTID	AAID	<u>Length</u>	Length			
5117337_f2_3	1497	6719	150 4	53	201	4.4e-16	l
Protein name				Locus	Name	Acc#	
				sp:YHC	HAEIN	P44583	
Description							
HYPOTHETICAL PRO	TEIN HI0227						
			ATT.	7.7			_
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
781932_£3_7	1498	6720	698	2094	543	3.1e-53	
Protein name				Locus	Name	Acc#	
115K outer membr protein	ane protein p	recursor	SusC	pir:JC	5027	JC6027	
Description							
			NT	AA	Score	Probability	_
ORF Name	NTID	AAID	Length	Length	SCOLE	FIODADIIICY	
11209542 <u></u> c3 <u></u> 31	1499	6721	73	222			
11209542G331 Protein name	1499	6721	73		s Name	Acc#	
	1499	6721	73		s Name	Acc#	
Protein name	[1499	6721	73 [2		s Name	Acc#	
Protein name  Description	[1499	6721		Locus	s Name		
Protein name  Description		6721 AAID	NT Length		Score	Acc# Probability	
Protein name  Description  NO-HIT	NTID		NT Length	Locus			
Protein name  Description  NO-HIT  ORF Name	NTID	AAID	NT Length	Locus  AA  Length	Score	Probability	
Protein name  Description  NO-HIT  ORF Name  24023442_c3_32	NTID 	AAID 6722	NT Length	Locus  AA  Length	Score 630 s Name	Probability	

ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Length	Score P	robability
25433212_£1_2	1501 6723	116 351	193	5.1e-14
Protein name		Loc	us Name	Acc#
alpha-N-acetylgluco	saminidase	gp:NT	A18209	Y18209
Description				
Nicotiana tabacum m	nRNA for alpha-N-ace	etylglucosamini	dase.	
		NT AA	Score P	robability
ORF Name	NTID AAID	Length Lengtl	<u> </u>	•
26681533_c2_26	1502 6724	202 609	327	3.5e-29
Protein name		Loc	us Name	Acc#
probable cationic a	amino acid transport	er pir:1	34694	T34694
Description				
		NT AA		
ORF Name	NTID AAID	Length Lengt	h Score F	Probability
29931309 <u>c1</u> 18	. 1503 6725	1299	195	1.9e-12
Protein name		Loc	cus Name	Acc#
immunoreactive 52kl	) antigen PG41	gp:Al	7175716	AF175716
Description				
Porphyromonas ging complete cds.	ivalis strain W50 i	mmunoreactive 5	2kD antigen	PG41 gene,
complete cus.		·		
ORF Name	NTID AAID	<u>NT AA</u> Length Lengt	h Score I	Probability
3.0.51.6.4.4.2 <u></u> .£1 <u></u> 1	1504 6726	1341	618	2.9e-60
Protein name		Loc	cus Name	Acc#
		sp:A	NAG_HUMAN	P54802
Description			n =	÷
GLUCOSAMINIDASE) (	NAG)			

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
10580062_c2_93	1505	6727	301	903	279	8.0e-24
Protein name				Locus	Name	Acc#
60kDa protein				gp:AB00	)4560	AB004560
Description						
Porphyromonas gi	ngivalis DNA	for 60kD	a proteir	n, comple	ete cds.	
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
13175950_c2_91	1506	6728	72	219		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT					i.	
			NT	AA		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 13869000c278		<u>AAID</u>	Length		Score 886	Probability
0			Length	Length	*	
13869000278  Protein name  115K outer membr	 	6729	Length	Length	886 s Name	1.6e-109
Protein name  115K outer membr	 	6729	Length	Length  Locus	886 s Name	1.6e-109 <u>Acc#</u>
13869000278  Protein name  115K outer membr	 	6729	Length	Length  Locus	886 s Name	1.6e-109 <u>Acc#</u>
Protein name  115K outer membr	 	6729	Length  999  SusC	Length  Locus  Dir:JC	886 s Name 6027	1.6e-109 Acc# JC6027
Protein name  115K outer membr	 	6729	Length  999  SusC	Length  Locus  pir:JC	886 s Name	1.6e-109 <u>Acc#</u>
Protein name  115K outer membra protein  Description	ntiD	6729 precursor	Length  999  SusC  NT  Length	Length  Locus  Dir:JC	886 s Name 6027	1.6e-109 Acc# JC6027
Protein name  115K outer membra protein  Description  ORF Name	ntiD	precursor  AAID	Length  999  SusC  NT  Length	Length  Locus  Locus  Dir:JC  AA  Length	886 s Name 6027	1.6e-109 Acc# JC6027
Protein name  115K outer membra protein  Description  ORF Name	ntiD	precursor  AAID	Length  999  SusC  NT  Length	Length  Locus  Locus  Dir:JC  AA  Length	886 S Name 6027 Score	Acc# JC6027  Probability

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
14647808_f2_20	1509	6731	287 8	64	126	4.3e-06	
Protein name				Locus	Name	Acc#	
				sp:YDI	ECOLI	P77402	
Description							
HYPOTHETICAL TRAN	SCRIPTIONAL	REGULATO	R IN AROL	-PPS INT	ERGENIC	REGION	
			NT	AA			
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability	
14660892_c1_74	1510	6732	207	24			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA		_ 1 1 1 2 1 .	
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
ORF Name 15835062_f1_18		<u>AAID</u> 6733	Length		Score 78	Probability 0.0048	]
1			Length	Length			]
15835062 <u></u> £1 <u></u> 18	1511		Length	Length	78 Name	0.0048	]
15835062_f1_18 Protein name	1511		Length	Length 291 Locus	78 Name	0.0048 <u>Acc#</u>	]
Protein name hypothetical prot	1511		Length	Length  291  Locus  pir:S7	78 Name	0.0048 <u>Acc#</u> S75372	]
Protein name hypothetical prot	1511		Length	Length 291 Locus	78 Name	0.0048 <u>Acc#</u>	]
Protein name  hypothetical prot  Description	1511 cein c04005	6733	Length  96  NT Length	Length  Locus  pir:S7	78 S Name	0.0048 <u>Acc#</u> S75372	]
Protein name  hypothetical prot  Description  ORF Name	1511 cein c04005	6733 AAID	Length  96  NT Length	Length Locus  Dir:S7  AA Length	78 S Name 5372 Score	0.0048 <u>Acc#</u> S75372  Probability	]
Protein name  hypothetical prot  Description  ORF Name	1511 cein c04005	6733 AAID	Length  96  NT Length	Length Locus  Dir:S7  AA Length	78 S Name Score 195 S Name	0.0048  Acc#  S75372  Probability  1.7e-12	]
Protein name  hypothetical prot  Description  ORF Name  20007812_c3_102	1511 cein c04005	6733 AAID	Length  96  NT Length	Length Locus  AA Length Locus  Locus  Locus  Locus  Locus  Locus	78 S Name Score 195 S Name	0.0048  Acc#  S75372  Probability  1.7e-12  Acc#	]

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
22683287_f3_56	1513	6735	486	.461			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT	-						
			NT	<u>AA</u>	Score	Probability	
ORF Name	NTID	AAID	Length	Length			
2449217.7 <u></u> c2 <u></u> 88	1514	6736	342	.029		×0	
Protein name				Locus	Name	Acc#	
Description							
NO-HIT	*	*		1			
		<del></del>	NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length	Score	FICHABILITEY	
24895165 <u>_</u> £1 <u>_</u> 13	1515	6737	65	L98	47	0.029	
Protein name				Locus	s Name	Acc#	
hypothetical prote	ein T11B7.2			pir:T24	1826	T24826	
Description							
	×	<u> </u>	NT	AA		Deckahiliter	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
26771041_c3_94	1516	6738	260	783	417	5.7e-39	
Protein name				Locus	s Name	Acc#	
hypothetical prot	ein C33G8.2			pir:T34	4137	T34137	
Description							
Y &		14.	NT	AA			
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability	
33772811 <u></u> f3 <u></u> 47	1517	6739	109	330			
Protein name				Locus	s Name	Acc#	
Description						,	
NO-HIT							

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
34510418_c1_63	1518 67	40	259 7	80	424	1.0e-39
Protein name				Locus	Name	Acc#
hypothetical protes	in F36H12.3			pir:T33	457	T33457
Description						
			NT	AA		- 1 7 '7 '1
ORF Name	NTID	AAID	Length	Length	Score	Probability
35734500 <u></u> c2 <u></u> 79	. 1519 67	41	512 1	539	202	7.9e-13
Protein name				Locus	Name	Acc#
unknown				gp:U967	71	U96771
Description						
Prevotella bryanti	i putative po	lygalac	turonase	,B-1,4-∈	ndogluca	nase, and
mannanase genes, co	mpiete cas; a	and dini		•		
mannanase genes, co	mpiete cus; a		NT	AA	Score	Probability
ORF Name	NTID	AAID			Score	Probability
	NTID		NT Length	AA	Score 430	Probability  2.4e-40
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length		*
ORF Name 3.6.1.9.8.9.5.8	<u>NTID</u> 1520 67	AAID	<u>NT</u> Length	AA Length	430 Name	2.4e-40
ORF Name  3.6.1.9.8.9.5.8	<u>NTID</u> 1520 67	AAID	<u>NT</u> Length	AA Length 37	430 Name	2.4e-40 <u>Acc#</u>
ORF Name  3.6.19.8.9.5.8	<u>NTID</u> 1520 67	AAID	NT Length 278 8	AA Length  37  Locus pir:T34	430 Name	2.4e-40 Acc# T34137
ORF Name  3.6.19.8.9.5.8	<u>NTID</u> 1520 67	AAID	<u>NT</u> Length	AA Length 37	430 Name	2.4e-40 <u>Acc#</u>
ORF Name  3.6.198.958	NTID  1520 67  in C33G8.2	AAID 742	NT Length 278 8	AA Length  37  Locus pir:T34	430 S Name	2.4e-40 Acc# T34137
ORF Name  3.6.198.958c276  Protein name  hypothetical prote  Description  ORF Name	NTID  1520 67  in C33G8.2	AAID AAID	NT Length 278 8	AA Length  Locus  pir:T34  AA Length  Length	430 S Name 1137 Score	2.4e-40  Acc#  T34137  Probability
ORF Name  3.6.198.958c276	NTID  1520 67  in C33G8.2	AAID AAID	NT Length 278 8	AA Length  Locus  pir:T34  AA Length  Length  Locus	430 S Name 1137 Score 545	2.4e-40  Acc#  T34137  Probability  3.1e-59
ORF Name  3.6.198.958c276	NTID  1520 67  in C33G8.2	AAID AAID	NT Length 278 8	AA Length  Locus  pir:T34  AA Length  Length  Locus	430 S Name 1137 Score 545 S Name	2.4e-40  Acc#  T34137  Probability  3.1e-59  Acc#

ORF Name	NTID AAID	<u>NT</u> <u>AA</u> Length Lengt	Score	Probability	
397175_c2_77	1522 6744	176   531	7		
Protein name		Loc	us Name	Acc#	
Description					
NO-HIT					
ORF Name	NTID AAID	<u>NT AA</u> Length Lengt	<u>Score</u>	Probability	
40.62906 <u></u> c3 <u></u> 101	1523 6745	785 2358	130	1.3e-08	
Protein name		Loc	us Name	Acc#	
		sp:F	'UA_YEREN	P46360	
Description					
PESTICIN RECEPTO	R PRECURSOR (IRPC)	(IPR65)			
		NT AA	Score	Probability	
ORF Name	NTID AAII		h Score	Probability	
ORF Name 4969091c397			h Score	Probability	
		Length Lengt	h Score	Probability  Acc#	
4969091 <u></u> c3 <u></u> 97		Length Lengt	<u>h</u>		
4969091c397 Protein name		Length Lengt	<u>h</u>		
Protein name  Description		Length Lengt	<u>h</u>	Acc#	
Protein name  Description		Length Lengt  138 417  Loc  NT AA	cus Name		
Protein name  Description  NO-HIT	1524 6746	Length Lengt  138 417  Loc  NT AA	cus Name	Acc#	
Protein name  Description  NO-HIT  ORF Name	1524 6746	Length Length  Loc  NT AA  Length Length  177 534	h Score	Acc# Probability	
Protein name  Description  NO-HIT  ORF Name  5900377cl66	1524 6746	Length Length  138 417  Loc  NT AA  Length Length  177 534	n Score	Acc#  Probability  5.6e-32	
Protein name  Description  NO-HIT  ORF Name  5900377cl66	1524 6746	Length Length  138 417  Loc  NT AA  Length Length  177 534	h Score  351  Cus Name	Acc#  Probability  5.6e-32  Acc#	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
	1526	6748		79		
6485055_c2_89	1526	0740				70
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA	_	D1 -1-1114
ORF Name	NTID	AAID	Length	Length	Score	Probability
103287f399	1527	6749	501 1	.506	479	1.5e-45
Protein name				Locus	Name	Acc#
immunoreactive 51kD	antigen l	PG52		gp:AF17	5719	AF175719
Description						
Porphyromonas gingi	valis str	ain W50 i	nmunoreac	ctive 51k	D antige	enPG52 gene,
complete cds.						(3)
			NT	AA	Cana	Drobability
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
ORF Name 11976562_c3_189	<u>NTID</u>	<u>AAID</u> 6750	Length		Score	Probability
		*	Length	Length	Score S Name	Probability  Acc#
1197.6562c3189		*	Length	Length		-
11976562c3189 Protein name		*	Length	Length		-
Protein name  Description		*	Length 61 [	Length  Locus		-
Protein name  Description		*	Length	Length		-
Protein name  Description  NO-HIT	1528	6750	Length  NT Length	Length  Locus	3 Name	Acc#
Protein name  Description  NO-HIT  ORF Name	NTID	6750	Length  NT Length	Length  Locus  AA  Length	Score	Acc# Probability
Protein name  Description  NO-HIT  ORF Name	NTID	6750	Length  NT Length	Length  Locus  AA  Length  Locus  Locus	Score	Acc# Probability  3.1e-13
Protein name  Description  NO-HIT  ORF Name	NTID	6750	Length  NT Length	Length  Locus  AA  Length  Locus  Locus	Score 139 Name	Acc#  Probability  3.1e-13  Acc#

_		NT AA	
ORF Name	NTID AAID	Length Length Score	Probability
13704552_c1_129	1530 6752	496 1491 1403	1.9e-143
Protein name		Locus Name	Acc#
		sp:6PGD_TREPA	083351
Description			
6-PHOSPHOGLUCONATE	DEHYDROGENASE, DEC	CARBOXYLATING,	
ORF Name	NTID AAID	NT AA Score Length Length	Probability
13758530_c3_190	1531 6753	136 411	
Protein name	J	Locus Name	Acc#
Description			
*			
NO-HIT			*
		NT AA G	
ORF Name	NTID AAID	Length Length Score	Probability
13907312 <u></u> c3 <u></u> 186	. 1532 6754	74 225 77	0.0096
Protein name		Locus Name	Acc#
putative signal tr	ansduction protein	GarA gp:AF173844	AF173844
Description			
Mycobacterium smeg	matis garA-contain:	ing gene cluster, partial	sequence.
ORF Name	NTID AAID	NT AA Score	Probability
13962757 <u></u> .a2 <u></u> .175	1533 6755	383 1152 350	7.2e-32
Protein name		Locus Name	Acc#
cytochrome d oxida	se subunit II	gp:AF001503	AF001503
Description	*		<u>.</u>
Calmonalla timbimu	rium cytochrome d	oxidase subunit I (cydA)	andcytochrome
d oxidase subunit I	I (cydB) genes, co	mplete cds.	

			NT	AA Tongth	Score	Probability
ORF Name	NTID	AAID	Length	Length	-	1 10 030
1444627_c1_132	1534 6	756	62 1	89	58	0.039
Protein name				Locus	Name	Acc#
ribosomal protein	S5			gp:U871	.45	U87145
Description						
Toxoplasma gondii	chloroplast,	complet	e genome			di
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
16441305_f3_103	1535	5757	236	11	244	1.2e-20
Protein name	141	*		Locus	Name	Acc#
hypothetical prote	ein b2381			pir:B6	5012	B65012
Description						÷ .
						*
				7. 7.		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 16517156£3101		<u>AAID</u> 6758	Length		Score	Probability  9.2e-71
			Length	Length		* *
16517156 <u></u> f3 <u></u> 101	1536		Length	Length	717 s Name	9.2e-71
16517156f3101 Protein name	1536		Length	Length  Locus	717 s Name	9.2e-71 <u>Acc#</u>
Protein name hypothetical protein	1536		Length 450	Length  1353  Locus  pir:S7	717 s Name	9.2e-71 <u>Acc#</u>
Protein name hypothetical protein	1536		Length	Length  Locus	717 s Name	9.2e-71 <u>Acc#</u>
Protein name hypothetical protein Description	1536   E	6758	Length  450  NT Length	Length  Locus  Locus  AA	717 5 Name 6946	9.2e-71 Acc# S76946
Protein name  hypothetical protein  Description  ORF Name	1536   E	AAID	Length  450  NT Length	Length Locus Locus AA Length	717 S Name 6946 Score	9.2e-71 <u>Acc#</u> S76946  Probability
Protein name  hypothetical protein  Description  ORF Name	1536   E	AAID	Length  450  NT Length	Length Locus Locus AA Length	717 S Name 6946 Score 63 S Name	9.2e-71 Acc# S76946  Probability  0.0078
Protein name  hypothetical protein  Description  ORF Name  194128_cl_130  Protein name	1536   E	AAID	Length  450  NT Length	Length  Locus  Locus  AA  Length  Locus  Locus  Locus  Locus  Locus	717 S Name 6946 Score 63 S Name	9.2e-71  Acc#  S76946  Probability  0.0078  Acc#

ORF Name	NTID AAI	NT D Length	AA Score	<u>Probability</u>
19687836 f3 87	1538   6760		 L290   945	6.4e-95
Protein name			Locus Name	Acc#
Procein name			sp:YCAJ_HAEIN	P45262
Description			<u> </u>	
HYPOTHETICAL PROT	rein Hi1590			
111101111111111111111111111111111111111		NET	7. 7.	
ORF Name	NTID AA	NT ID Length	<u>AA</u> <u>Length</u> <u>Score</u>	Probability
2068766_c1_143	1539 6761	525	1578 1098	3.9e-111
Protein name			Locus Name	Acc#
			sp:CYDA_AZOVI	Q09049
Description				
CYTOCHROME D UBI	QUINOL OXIDASE SU	BUNIT I,		
		NT	AA Score	Probability
ORF Name	NTID AA	NT	AA Length Score	Probability
ORF Name 20.73.58.78216.3	NTID AA	<u>NT</u> ID Length	— Score	Probability
	NTID AA	<u>NT</u> ID Length	Length Score	Probability  Acc#
20.73.58.78 <u></u> c2 <u></u> 16.3	NTID AA	<u>NT</u> ID Length	Length Score	
20735878_c2_163 Protein name	NTID AA	<u>NT</u> ID Length	Length Score	
Protein name  Description	NTID AA	ID Length	Length Score Locus Name	Acc#
Protein name  Description	NTID AA	NT Length  142	Length  Locus Name	
20735878c2163  Protein name  Description  NO-HIT	NTID AA 1540 6762	NT Length  142  NT Length  Length	Length  Locus Name  AA Score	Acc#
20735878c2163  Protein name  Description  NO-HIT  ORF Name	NTID AA 1540 6762	NT Length  142  NT Length  Length	Length  Locus Name  AA Length  Score	Acc# Probability
20.73.58.78	NTID AA 1540 6762	NT Length  142  NT Length  Length	Length  Locus Name  Locus Name  AA Length Score  1341  911	Acc# Probability  2.6e-91
Protein name  Description  NO-HIT  ORF Name  220.43.77t270  Protein name  RumB(R391)  Description	NTID AA 1540 6762	NT Length  142  NT Length  ANT Length  446	Length  Locus Name  Locus Name  AA Length  Score  1341  Locus Name  gp:XXU13633	Acc# Probability  2.6e-91 Acc# U13633

		* * T.D.	NT Tangth	<u>AA</u> Length	Score	Probability
ORF Name	NTID	AAID	Length		0.55	1.7e-23
22273312_c1_131	1542	6764	287 8	64	271	
Protein name				Locus	s Name	Acc#
urea transport pro	tein			gp:AF1	67577	AF167577
Description						
Actinobacillus ple partial cds; and pu cytoplasmic membrar homolog (cbiM), col transport ATP-bind	utative peri ne protein (d palt membran	plasmic cbiL), co e transpo	binding p bbalt mem ort prote	rotein branetra inhomolo	(cbiK),pu ansport p og (cbiQ)	orotein , cobalt
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
22535925 <u></u> c2 <u>.</u> 176	1543	6765	359	.080	184	2.2e-12
Protein name				Locu	s Name	Acc#
molybdate metabol:	sm regulato	r		pir:B6	4979	B64979
				1 1		
Description					.4	<u>,                                    </u>
Description	6		NT	AA	Score	Probability
Description ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
		<u>AAID</u> 6766	Length		Score 651	Probability  9.1e-64
ORF Name			Length	Length		1
ORF Name 22941306t262	1544	6766	Length	Length	651 s Name	9.1e-64
ORF Name 22941306_f2_62	1544	6766	Length	Length  777  Locu	651 s Name	9.1e-64 <u>Acc#</u>
ORF Name  229413.06_f2_62  Protein name  ABC transporter,	1544	6766	Length 258	Length 777  Locu pir:H7	651 s Name	9.1e-64 <u>Acc#</u> H72385
ORF Name  229413.06_f2_62  Protein name  ABC transporter,	1544	6766	Length	Length  777  Locu	651 s Name	9.1e-64 <u>Acc#</u>
ORF Name  229413.06f262  Protein name  ABC transporter,  Description	1544 ATP-binding NTID	6766	Length  258  NT  Length	Length  Locu  pir:H7	651 s Name 2385	9.1e-64 <u>Acc#</u> H72385
ORF Name  229413.06_f2_62  Protein name  ABC transporter,  Description  ORF Name	1544 ATP-binding NTID	protein  AAID	Length  258  NT  Length	Length  Locu  Dir:H7  AA  Length	651 s Name 2385	9.1e-64 <u>Acc#</u> H72385
ORF Name  229413.06f262  Protein name  ABC transporter,  Description  ORF Name  23985880f126	1544 ATP-binding NTID	protein  AAID	Length  258  NT  Length	Length  Locu  Dir:H7  AA  Length	Score	9.1e-64  Acc#  H72385  Probability

			$\underline{ ext{NT}}$	$\underline{\underline{AA}}$	Score	Probability	
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>			_
24100265_c3_188	1546	6768	505 1	518	1300	1.5e-132	
Protein name				Locus	Name	Acc#	
				sp:G6PI	_ACTAC	P77809	
Description							
GLUCOSE-6-PHOSPHATE	E 1-DEHYDR	OGENASE,	(G6PD)				
							_
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
24229800 f3 98	1547	6769	79	240			
Protein name	J L			Locus	s Name	Acc#	
Description							
NO-HIT					- 1	9	
			NT	AA	~	Declaration des	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 24647180_c3_191		<u>AAID</u> 6770	Length		Score	Probability 0.0017	_
		<del></del>	Length	Length			]
24647180c3191	. 1548	<del></del>	Length	Length	101 s Name	0.0017	]
24647180_c3_191	. 1548	<del></del>	Length	Length 2073 Locus	101 s Name	0.0017 <u>Acc#</u>	]
Protein name  hypothetical prote	. 1548	<del></del>	Length	Length 2073 Locus	101 s Name	0.0017 <u>Acc#</u>	
Protein name  hypothetical prote	. 1548	<del></del>	Length	Length 2073 Locus	101 s Name	0.0017 <u>Acc#</u>	]
Protein name  hypothetical protein Description	1548 in MTH357	6770	Length 690 NT Length	Length  2073  Locu  pir:A6	101 s Name 9146	0.0017 Acc# A69146	]
Protein name  hypothetical protein Description  ORF Name	1548 in MTH357	6770 AAID	Length 690 NT Length	Length  Locu  Locu  pir:A6  AA  Length  894	101 s Name 9146 Score	0.0017  Acc#  A69146  Probability	]
Protein name  hypothetical protein  Description  ORF Name	. 1548 in MTH357  NTID . 1549	AAID	Length 690 NT Length	Length  Locu  Locu  pir:A6  AA  Length  894	101 s Name 9146 Score 392 s Name	0.0017  Acc#  A69146  Probability  3.8e-35	]
Protein name  hypothetical protein Description  ORF Name  25517013_f1_1  Protein name	. 1548 in MTH357  NTID . 1549	AAID	Length 690 NT Length	Length  Locus  Locus  Pir:A6  AA  Length  B94  Locus	101 s Name 9146 Score 392 s Name	0.0017  Acc#  A69146  Probability  3.8e-35  Acc#	

			$\underline{ ext{NT}}$	AA	Score	Probability	
ORF Name	NTID	AAID	Length	Length			
25667675_c3_205	1550 6	772	341 1	1026			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	<u>Length</u>			
25970016f3100	1551 6	773	411	1236	326	2.5e-29	
Protein name				Locus	Name	Acc#	
probable membrane	protein b0878	3		pir:F64	826	F64826	
Description							
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
31769537 <u>_</u> £1 <u>_</u> 24	1552 6	774	206	621	222	2.0e-17	
Protein name	*	٠.		Locus	Name	Acc#	
				sp:YEH	J_ECOLI		
Description							
HYPOTHETICAL 62.1	KD PROTEIN I	N MOLR-I	BGLX INT	ERGENIC I	REGION P	RECURSOR	
			NT	AA		D11111	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
318863082159	1553 6	775	259	780	387	8.6e-36	
Protein name				Locu	s Name	Acc#	
probable glucose-	6-phosphate 1	-dehydro	genase	pir:C7	1319	C71319	
Description							
			NT	AA			<u></u>
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability	
3257635 <u></u> c1134	1554 6	776	426	1281			
Protein name				Locu	s Name	Acc#	
Description							
NO-HIT							]

			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length		
34382687_c3_193	1555	6777	419	1260		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length		
3945257 <u></u> £3 <u></u> 102	1556	6778	158	477		·
Protein name				Locus	s Name	Acc#
Description						
NO-HIT						γ.
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	SCOLE	FIODADITICY
4103890 <u></u> c1147	1557	6779	515	1548	145	9.4e-08
Protein name				Locu	s Name	Acc#
conserved hypot	hetical protei	n AF0444		pir:D6	9305	D69305
Description						
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	BCOTC	11020211107
4737662 <u></u> £1 <u></u> 3.9	1558	6780	395	1188	578	4.9e-56
Protein name				Locu	s Name	Acc#
probable glutam	ate/ aspartate	e transpo	rter	pir:G7	1309	G71309
Description						
			NT	AA		
ORF Name	NTID	AAID	Length		Score	Probability
5117762 <u></u> .£3 <u></u> .106	1559	6781	149	450	304	5.4e-27
Protein name				Locu	s Name	Acc#
RumA(R391)				gp:XXU	13633	U13633
Description						
IncJ plasmid R	391 rumA(R391)	and rum	3(R391) S	genes, co	mplete o	cds.

μ,	na.
**	1000
l,	,221
Ţ.	ngil
7	7
4	
	Y225
ij;	ses:
ş	22,
4	22
gf:	1 1
ij.	HANT TOTAL
2	THE PERSON
ŧ,	ozii
	117 127 137
ž.	. 5
	Her
**	
**	
The Street	
The State of	4.15 H H 4.15
海 本 海 本 海 本 海	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
海 本 海 本 海 本 海	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
報 等 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報	
報 等 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報	
報 等 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報	
報 等 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報 報	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

		7 7 T.D	NT	<u>AA</u> Length	Score	Probability	
ORF Name	NTID	AAID	Length		IEE/	1.7e-53	
5275250_£2_47	1560	6782	319	060	554		
Protein name				Locus	Name	Acc#	
				sp:DHGY	METEX	Q59516	
Description							
REDUCTASE) (HPR-A)							
		-	NT	AA	C.c.	Probability	_
ORF Name	NTID	AAID	Length	Length	Score	PIODADITICY	
7287787_c1_133	1561	6783	267	304			
Protein name				Locus	Name_	Acc#	
Description							
•							
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
	<u>NTID</u>	<u>AAID</u>	Length		Score	Probability  5.1e-09	]
9770137_c1_145		<del></del>	Length	Length	,		]
		<del></del>	Length	Length	109 S Name	5.1e-09	]
97.7.0137_c1_145		<del></del>	Length	Length 864 Locus	109 S Name	5.1e-09	]
Protein name  Description	1562	6784	Length 287	Length  Locus  gp:AB0	109 S Name	5.1e-09 Acc#	
97.7.0137_c1_145	1562	6784	Length  287  -SAKURA,	Length  Locus  gp:AB0	109 S Name	5.1e-09 Acc#	
Protein name  Description	1562	6784	Length 287	Length  Locus  gp:AB0	109 S Name	5.1e-09 Acc#	]
Protein name  Description  Agrobacterium tume	1562 faciens p	6784 Lasmid pTi	Length  287  -SAKURA,	Length  Locus  gp:AB0  complete	109 S Name 16260 S sequen	5.1e-09 Acc#	
Protein name  Description  Agrobacterium tume  ORF Name	1562 faciens p	6784 Lasmid pTi	Length  287  -SAKURA,  NT  Length	Length  Locus  gp:AB0  complete  AA  Length	109 S Name 16260 E sequen Score	5.1e-09 Acc#  ce.  Probability	]
Protein name  Description  Agrobacterium tume  ORF Name  9922057_c3_206	faciens p.  NTID 1563	6784 Lasmid pTi AAID	Length  287  -SAKURA,  NT  Length	Length  Locus  gp:AB0  complete  AA  Length	109 S Name 16260 Score 293 S Name	5.1e-09 Acc#  ce.  Probability  8.0e-26	
Protein name  Description  Agrobacterium tume  ORF Name  9922057_c3_206	faciens p.  NTID 1563	6784 Lasmid pTi AAID	Length  287  -SAKURA,  NT  Length	Length  Locus  gp:AB0  complete  AA  Length  1299  Locu	109 S Name 16260 Score 293 S Name	5.1e-09 Acc#  ce.  Probability  8.0e-26 Acc#	]

ORF Name	NTID AAID	NT AA Score Probability Length Length
13869003_f3_21	1564 6786	535 1608 123 0.0010
Protein name		Locus Name Acc#
glycoprotein Vp260	-like protein A18L	pir:T17508 T17508
Description		
		NT AA Score Probability
ORF Name	NTID AAID	Length Length
23864381 <u></u> c2 <u></u> 38	1565 6787	474 1425 747 6.1e-74
Protein name		Locus Name Acc#
metabolite transpo	rt protein homolog	ywtG pir:E70070 E70070
Description		
		NT AA Score Probability
ORF Name	NTID AAID	Length Length
25632943 <u></u> .c1 <u></u> 29	1566 6788	184 555
Protein name		Locus Name Acc#
Description		
NO-HIT		
		NT AA Score Probability
ORF Name	NTID AAID	Length Length
26364040 <u></u> f1 <u></u> 6	1567 6789	[186]
Protein name		Locus Name Acc#
Description		
NO-HIT		
		NT AA Score Probability
ORF Name	NTID AAID	Length Length
3.340.026.0 <u></u> £3 <u></u> 23	1568 6790	
Protein name	-	Locus Name Acc#
STARP antigen		gp:PFSTARP Z26314
Description		
P.falciparum gene	for STARP antigen.	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
791406_c3_44	1569 6	791		213	77	0.026	
Protein name			L	Locus	Name	Acc#	_
				sp:ATP6	_ACACA	Q37385	
Description	ei)						
ATP SYNTHASE A C	HAIN, (PROTEIN	6)					
			NT	AA	Score	Probability	
ORF Name	NTID	AAID	Length	<u>Length</u>	-		
9862501_c3_41	1570	792	109	330		<b>7</b>	
Protein name				Locus	s Name	Acc#	
Description							
NO-HIT					•		
		7.7.7	NT	<u>AA</u> Length	Score	Probability	
ORF Name	NTID	<u>AAID</u> 793	Length 205	618 T	545	1.6e-52	7
14664052 <u></u> £3 <u></u> .7	1571 6	/93	203		s Name	Acc#	
<u>Protein name</u>				gp:PGP		X95938	
Description				90.101			
P.gingivalis rnh	B & pgaA genes	& orfs	150, 19	7, 202 &	199.		1
1.9111911			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability	
34406512 <u></u> £1 <u></u> 4	1572 6	794	311	936	663	4.9e-65	
Protein name				Locu	s Name	Acc#	
2,3-bisphosphogl	ycerate-indepe	ndent		gp:AF1	20090	AF120090	
Description						· .	_
Bacillus megater mutase (pgm) gene	rium 2,3-bispho	sphogly	cerate-i	ndepende	ntphosph	oglycerate	
	,						J

ş.	#### ###
***	1107
	127
q	11211
î	Henry
1	: #
	Dist.
100	11117
100	1
ķ	100
ST-west	Genous .
2	iπβ
į.	pri
ij,	100
į.	1124;
ħ	1134
翻	
ij.	135
355	ng?
	19845
44	1000
2	
il,	æ
	me
Ę	222
£)	201
ij	
di,	210 AU
6	AU

			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	Length		
36135311_c1_9	1573	6795	315 9	48	444	7.8e-42
Protein name				Locus	Name	Acc#
probable transport	protein			pir:A75	272	A75272
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
3.6.3.3.0.1.7.5 <u></u> £1 <u></u> 5	1574	6796	62	L86		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	AA		_ 1 1 1 2 1 1 1 1
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
1503.9156 <u></u> ±3 <u></u> 6	1575	6797	192	579	398	4.7e-36
Protein name				Locus	s Name	Acc#
Protein name  putative large sec	creted prote	ein		Locus gp:SCF		Acc# AL117669
	creted prote	ein				
putative large sec		- 8				
putative large sec		- 8	NT	gp:SCF		AL117669
putative large sec		- 8	NT Length			
putative large second secretary putative large second secretary second s	icolor cosm	id F12.	Length	gp:SCF	12	AL117669
putative large second second putative large second	icolor cosm	id F12.	Length	gp:SCF	Score	AL117669  Probability
putative large second secretary putative large second secretary second s	icolor cosm	id F12.	Length	gp:SCF	Score 79 s Name	AL117669  Probability  0.042
putative large second second putative large second	icolor cosm  NTID	id F12. <u>AAID</u> 6798	Length	AA Length 270 Locu	Score 79 s Name	AL117669  Probability  0.042

			$\overline{ ext{NT}}$	$\underline{AA}$	Score	Probability
ORF Name	NTID	AAID	Length	Length		
24343756_c1_7	1577	6799	62	189		
Protein name				Locus	Name	Acc#
Danishion						
Description						
NO-HIT						
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	<u>acore</u>	FIODADITICY
4860650 <u>_</u> t1 <u>_</u> 3	1578	6800	95	288	75	0.0099
Protein name		<del>-</del>		Locus	Name	Acc#
ct602 hypothetical	protein			pir:F72	2036	F72036
Description						
		7 7 TI	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name	NTID	AAID				
10660763_c3_339	1579	6801	387	1164		
Protein name				Locu	s Name	Acc#
Description						
NO-HIT					*	
			NT	<u>AA</u>	Score	Probability
ORF Name	$\underline{\mathtt{NTID}}$	AAID	<u>Length</u>	Length		
1176302_f1_16	1580	6802	799	2400	150	3.0e-14
Protein name			,	Locu	s Name	Acc#
putative TonB-depe	endent out	er membrar	ie	gp:AF0	48749	AF048749
receptor						<del></del>
Description						*
Bacteroides fragi	lis capsul	ar polysa	ccharide	biosynth	esis ope	ron, complete
sequence.		_				

ORF Name	NTID AAID	NT AA So	ore Probability
12117076_c1_211	1581 6803	[67] [204]	
Protein name		Locus Na	ame Acc#
Description			
NO-HIT			
ORF Name	NTID AAID	NT AA So Length Length	core Probability
12506402 <u></u> £1 <u></u> 15	1582 6804	954 2865 24	9 4.1e-17
Protein name		Locus Na	Acc#
putative histidine	protein kinase	gp:REU825	U82564
Description			
hydrogenase-like p	protein small subuni !) gene,and putative	t(hoxB) gene, hydrod histidine protein l	genase-like protein kinase (hoxJ) gene,
complete cds, and ni	ckel permease (hoxN	) gene, partial cds	
complete cds, and ni	ckel permease (hoxN	) gene, partial cds	core Probability
complete cds,and ni	ckel permease (hoxN	) gene, partial cds  NT AA So	core Probability
Complete cds, and ni	ckel permease (hoxN	NT AA So Length Length	core Probability  1 4.9e-33
ORF Name  12540880_c3_343  Protein name  capsular polysacch	ckel permease (hoxN	NT AA Length Length 343 1032 36	core Probability  1 4.9e-33  ame Acc#
ORF Name  12540880_c3_343  Protein name  capsular polysacch yveT	1583 6805	NT AA Length Length 343 1032 36	core Probability  1 4.9e-33  ame Acc#
ORF Name  12540880_c3_343  Protein name  capsular polysacch	1583 6805	NT AA Length Length 343 1032 36	core Probability  1 4.9e-33  ame Acc#
ORF Name  125408803343  Protein name  capsular polysacch yveT  Description	NTID AAID 1583 6805 haride biosynthesis	NT AA Solution NT AA Length Length Length Locus Not the North North NT AA Solution NT AA Solutio	core Probability  1 4.9e-33  ame Acc#
ORF Name  12540880_c3_343  Protein name  capsular polysacch yveT  Description  ORF Name	NTID AAID  aride biosynthesis  NTID AAID  AAID  AAID  AAID	NT AA Solution AA Length Length Locus Not pir:A7003  NT AA Length Length Solution Not pir:A7003	core Probability  1 4.9e-33  ame Acc#  7 A70037
ORF Name  125408803343  Protein name  capsular polysacch yveT  Description  ORF Name  127140623354	NTID AAID  aride biosynthesis  NTID AAID  AAID  AAID  AAID	NT   AA   So	Probability  1 4.9e-33  ame Acc#  A70037  Core Probability
ORF Name  12540880_c3_343  Protein name  capsular polysacch yveT  Description  ORF Name	NTID AAID  aride biosynthesis  NTID AAID  AAID  AAID  AAID	NT AA Solution AA Length Length Locus Not pir:A7003  NT AA Length Length Solution Not pir:A7003	Probability  1 4.9e-33  ame Acc#  A70037  Core Probability
ORF Name  125408803343  Protein name  capsular polysacch yveT  Description  ORF Name  127140623354	NTID AAID  aride biosynthesis  NTID AAID  AAID  AAID  AAID	NT   AA   So	Probability  1 4.9e-33  ame Acc#  A70037  Core Probability

			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	Length		
1290933_f3_136	1585	6807	143 4	32	165	2.9e-12
Protein name				Locus	s Name	Acc#
hypothetical prote	in slr1861			pir:S7	7097	S77097
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
12923260c2271	. 1586	6808	516	.551	207	1.6e-13
Protein name				Locu	s Name	Acc#
putative flippase				gp:AF1	25164	AF125164
Description						
Bacteroides tragil	is 638R pc	lysacchar	ide B (PS	B2) bi	osynthes:	islocus,
_						
complete sequence;	and unknow	m genes.		14		
complete sequence;	and unknow	m genes.	NT	AA		
complete sequence; ORF Name	and unknow	AAID	NT Length	AA Length	Score	Probability
	NTID		Length		<u>Score</u>	Probability 2.5e-13
ORF Name	NTID	AAID	Length	Length		
ORF Name 14070180_c1_202	NTID [1587	AAID	Length	Length	175 s Name	2.5e-13
ORF Name  14070180c1202  Protein name	NTID [1587	AAID	Length	Length 453 Locu	175 s Name	2.5e-13 <u>Acc#</u>
ORF Name  14070180_c1_202  Protein name  hypothetical prote	NTID [1587	AAID	Length	Length  153  Locu  pir:S2	175 s Name	2.5e-13 Acc# S28678
ORF Name  14070180_c1_202  Protein name  hypothetical prote	NTID [1587	AAID	Length	Length 453 Locu	175 s Name	2.5e-13 <u>Acc#</u>
ORF Name  14070180_c1_202  Protein name  hypothetical prote  Description	NTID 1587 sin 1	<u>AAID</u> [6809	Length  150  NT Length	Length Locu Locu Pir:S2	175 s Name 8678	2.5e-13 Acc# S28678
ORF Name  140.70180c1202  Protein name  hypothetical prote  Description  ORF Name	NTID 1587 sin 1	<u>AAID</u> [6809	Length  150  NT  Length	Length Locu Locu Locu AA Length Locu Locu	175 s Name 8678 Score	2.5e-13 <u>Acc#</u> S28678  Probability
ORF Name  140.70.180c1202  Protein name  hypothetical prote  Description  ORF Name  1448442c1203	NTID 1587  Pin 1  NTID 1588	<u>AAID</u> 6809  AAID	Length  150  NT Length  354	Length Locu Locu Locu AA Length Locu Locu	175 s Name 8678 Score 599 as Name	2.5e-13 Acc# S28678  Probability  2.9e-58
ORF Name  140.70180c1202  Protein name  hypothetical protein  Description  ORF Name  1448442_c1203  Protein name	NTID 1587  Pin 1  NTID 1588	<u>AAID</u> 6809  AAID	Length  150  NT Length  354	Length Locu  Pir:S2  AA Length  1065 Locu	175 s Name 8678 Score 599 as Name	2.5e-13 Acc# S28678  Probability  2.9e-58 Acc#

ORF Name	- NTID AAID	<u>NT</u> <u>AA</u> Length Length <u>Score</u>	Probability
14567135_c1_201	1589 6811	369 1110 118	0.00040
Protein name		Locus Name	Acc#
immunoreactive 4	3kD antigen PG32	gp:AF175714	AF175714
Description			-
Porphyromonas gi complete cds.	ngivalis strain W50	immunoreactive 43kD antige	nPG32 gene,
ORF Name	NTID AAID	NT <u>AA</u> Score Length Length	Probability
14658342_f2_65	1590 6812	583 1752 139	2.5e-06
Protein name		Locus Name	Acc#
hypothetical pro	otein SPAC17G6.19c	pir:T37851	T37851
Description			
	*	NT AA Score	Probability
ORF Name	NTID AAID	Length Length Score	Probability
ORF Name		Length Length Score  105 318 161	7.6e-12
14726512_f1_11 Protein name	1591 6813	Length Length Score  105 318 161  Locus Name	7.6e-12 <u>Acc#</u>
14726512 <u></u> £1 <u></u> 11	1591 6813	Length Length Score  105 318 161	7.6e-12
14726512_f1_11 Protein name	1591 6813	Length Length Score  105 318 161  Locus Name	7.6e-12 <u>Acc#</u>
Protein name hypothetical pro	1591   6813 Stein sir1856	Length Length Score  105 318 161  Locus Name  pir:S77093	7.6e-12 <u>Acc#</u>
Protein name hypothetical pro	1591 6813 6813 6813 6813 6813 6813 6813 681	Length Length Score  105 318 161  Locus Name  pir:S77093  NT AA Length Length Score	7.6e-12  Acc#  S77093  Probability
Protein name  hypothetical production	1591 6813 6813 6813 6813 6813 6813 6813 681	Length         Length         Score           105         318         161           Locus         Name           pir:S77093           NT         AA           Length         Length           647         1944         979	7.6e-12  Acc#  S77093  Probability  1.6e-98
Protein name  hypothetical pro  Description  ORF Name	1591 6813 6813 6813 6813 6813 6813 6813 681	Length         Length         Score           105         318         161           Locus         Name           pir:S77093           NT         AA           Length         Length           647         1944           Pir:S77093	7.6e-12  Acc#  S77093  Probability  1.6e-98  Acc#
Protein name  hypothetical production  Description  ORF Name  15628390_f1_9  Protein name	1591 6813 6813 6813 6813 6813 6813 6813 681	Length         Length         Score           105         318         161           Locus         Name           pir:S77093           NT         AA           Length         Length           647         1944         979	7.6e-12  Acc#  S77093  Probability  1.6e-98
Protein name  hypothetical pro  Description  ORF Name	1591 6813 6813 6813 6813 6813 6813 6813 681	Length         Length         Score           105         318         161           Locus         Name           pir:S77093           NT         AA           Length         Length           647         1944           Pir:S77093	7.6e-12  Acc#  S77093  Probability  1.6e-98  Acc#

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
	1593	6815	383	152	322	6.6e-29
Protein name				Locus	Name	Acc#
Cps1K				gp:AF1	55804	AF155804
Description						
Streptococcus si (cps2F), Cps1G (cgenes, complete	cps1G), Cps1H	(cps1H),	Cps1I (	cpsll), a	anacpsio	cps1J)
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
15822807_f1_2 Protein name	1594	6816	549	Locus	s Name	Acc#
Description						
по-ніт						
NO-HIT ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
		<u>AAID</u> 6817	Length		Score	Probability  0.037
ORF Name  171902_f1_31  Protein name	1595	6817	Length	Length 204 Locu	49 s Name	0.037 <u>Acc#</u>
ORF Name 171902_f1_31	1595	6817	Length	Length 204 Locu	49 s Name	0.037
ORF Name 171902_f1_31  Protein name probable RNA-di	1595	6817	Length	Length 204 Locu	49 s Name	0.037 <u>Acc#</u>
ORF Name 171902_f1_31  Protein name probable RNA-ditranscriptase	1595	6817	Length	Length 204 Locu	49 s Name	0.037 <u>Acc#</u>
ORF Name  171902f131  Protein name  probable RNA-ditranscriptase  Description	rected DNA po	6817	Length 67 :reverse  NT Length	Length  Locu  pir:S2	49 s Name 0016	0.037 Acc# S20016
ORF Name  1.7.19.02f131  Protein name  probable RNA-ditranscriptase  Description  ORF Name	rected DNA po	6817  Tymerase,	Length 67 :reverse  NT Length	Length  Locu  pir:S2  AA  Length  1473	49 s Name 0016	0.037 Acc# S20016
ORF Name  1.719.02f131  Protein name  probable RNA-ditranscriptase  Description  ORF Name  19.710.937f2123	rected DNA po	6817  Tymerase,	Length 67 :reverse  NT Length	Length  Locu  pir:S2  AA  Length  1473	Mame 0016 Score	0.037 Acc# S20016 Probability

	440 840
*34	MAG
į,	
Separation of the last	145
	ten:
ΫH	1257
Hazzir	156 207
	Therese
£J	GG.
Chertife.	1100 1100 1100
100	7
ä	
Sept.	1177
4	arii
	1
3\$1	H
35	507
Sharks	1223
	93
35	ÇÇ.
-25	HEL.
P.A.	110
.53	
F.	
e.	100

			NT	$\underline{\underline{AA}}$	Score	Probability	
ORF Name	NTID	AAID	<u>Length</u>	Length			
19728412_c3_391	1597	6819	459 1	380	472	9.2e-54	
Protein name				Locus	Name	Acc#	
folylpolyglutamate	synthase/c	dihydrofol	.ate	pir:D72	411	D72411	
synthase							
Description							
		a a to	NT	<u>AA</u> Length	Score	Probability	
ORF Name	NTID	AAID	Length	<u>Herigeri</u>			
2032137 <u></u> c1 <u></u> 213	1598	6820	88 2	267			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
							•
			N.T.	7. 7.			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name 21651557c3338	NTID	<u>AAID</u>	Length		Score	Probability	
21651557 <u></u> c3 <u></u> 338			Length	Length			
21651557 <u>c3</u> 338 Protein name	1599		Length	Length	128 s Name	2.2e-05	
Protein name  hypothetical prote:	1599		Length	Length  Locus	128 s Name	2.2e-05 <u>Acc#</u>	
21651557 <u>c3</u> 338 Protein name	1599		Length	Length Locus	128 s Name	2.2e-05 <u>Acc#</u>	
Protein name  hypothetical prote:  Description	1599		Length	Length  Locus	128 s Name	2.2e-05 <u>Acc#</u>	
Protein name  hypothetical prote:	1599 in RP338	6821	Length  354  NT Length	Length Locus	128 s Name 1690	2.2e-05 Acc# D71690	
Protein name  hypothetical prote:  Description  ORF Name	1599 in RP338	6821 AAID	Length  354  NT Length	Length Locus pir:D7  AA Length	128 s Name 1690	2.2e-05 Acc# D71690	
Protein name hypothetical protein Description  ORF Name  23455077_c3_348	1599 in RP338	6821 AAID	Length  354  NT Length	Length Locus pir:D7  AA Length	128 s Name 1690 Score	2.2e-05 Acc# D71690 Probability	
Protein name  hypothetical prote:  Description  ORF Name  23455077c3348	1599 in RP338	6821 AAID	Length  354  NT Length	Length Locus pir:D7  AA Length	128 s Name 1690 Score	2.2e-05 Acc# D71690 Probability	

			NT	AA	Cacaa	December 11 decemb
ORF Name	NTID	AAID	Length	Length	Score	Probability
23594641_c1_195	1601	6823	250	753	316	2.9e-28
Protein name				Locus	Name	Acc#
putative UDP-N-actransferase	etyl-D-manno	samine		gp:SPU(	9239	U09239
Description				-		
Streptococcus pne operon, (cps19fABC cds.						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
23632802_£3_146	1602	6824	270	313	412	1.9e-38
Protein name				Locus	Name	Acc#
				gp:AB00	8550	AB008550
Description						
Pseudomonas aerug	inosa phage	phi CTX,	complete	genome	sequence	2.
			NT	AA		
ORF Name	NTID	AAID			Score	Probability
ORF Name 23.7.7.0.92£31.53	NTID		<u>NT</u> Length	AA		
	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
23.77.0.92 <u></u> £3 <u></u> 1.53	NTID 1603	AAID	<u>NT</u> Length	AA Length	Score 219 Name	Probability
2377092_f3_153 Protein name	NTID 1603	AAID	<u>NT</u> Length	AA Length 95 Locus	Score 219 Name	Probability  [1.3e-17  Acc#
2377092_t3_153	NTID 1603  nsferase lis 638R pol	AAID 6825 ysacchari	NT Length 64	AA Length  95  Locus  gp:AF12	Score 219 Name	Probability  1.3e-17  Acc#  AF125164
Protein name  putative aminotral  Description  Bacteroides fragi	NTID 1603  nsferase lis 638R pol	AAID 6825 ysacchari	NT Length 64	AA Length  95  Locus  gp:AF12	Score 219 Name 25164 synthesi	Probability  1.3e-17  Acc#  AF125164  Slocus,
Protein name  putative aminotral  Description  Bacteroides fragi	NTID 1603  nsferase lis 638R pol	AAID 6825 ysacchari	NT Length  64  1	AA Length  95 Locus  gp:AF12  B2) bio	Score 219 Name	Probability  1.3e-17  Acc#  AF125164
Protein name  Putative aminotral  Description  Bacteroides fragicomplete sequence;	NTID  1603  Insterase  Lis 638R polyand unknown  NTID	AAID 6825 ysacchari genes.	NT Length  64 1  de B (PS	AA Length  95  Locus  gp:AF12  B2) bio	Score 219 Name 25164 synthesi	Probability  1.3e-17  Acc#  AF125164  Slocus,
Protein name  Putative aminotral  Description  Bacteroides fragicomplete sequence;  ORF Name	NTID  1603  Insterase  Lis 638R polyand unknown  NTID	AAID 6825 ysacchari genes.  AAID	NT Length  MT Length	AA Length  95  Locus  gp:AF12  B2) bio  AA Length	Score  219  Name  5164  synthesi	Probability  1.3e-17  Acc#  AF125164  Slocus,  Probability
Protein name  Putative aminotral  Description  Bacteroides fragicomplete sequence;  ORF Name  24038512_f3_134	NTID  1603  Insterase  Lis 638R polyand unknown  NTID	AAID 6825 ysacchari genes.  AAID	NT Length  MT Length	AA Length  95  Locus  gp:AF12  B2) bio  AA Length	Score  219  Name  5164  synthesi  Score  375  Name	Probability  1.3e-17  Acc#  AF125164  Slocus,  Probability  1.6e-34
Protein name  Putative aminotral  Description  Bacteroides fragicomplete sequence;  ORF Name  24038512_f3_134	NTID  1603  Insterase  Lis 638R polyand unknown  NTID	AAID 6825 ysacchari genes.  AAID	NT Length  MT Length	AA Length  95  Locus  Gp:AF12  B2) bio  AA Length  56  Locus	Score  219  Name  5164  synthesi  Score  375  Name	Probability  1.3e-17  Acc#  AF125164  Slocus,  Probability  1.6e-34  Acc#

	NETE	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name	NTID	<del>,</del>					
24412537_f1_20	1605	6827	276	31		7 11	
Protein name				Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA		Dla-la-17-1	
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability	
24413887 <u></u> c1200	. 1606	6828	178	537	72	0.048	
Protein name				Locus	s Name	Acc#	
				sp:Y23	5_METJA	Q57687	
Description	,						
HYPOTHETICAL PROTE	IN MJ0235				н		1
			NT	AA			
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
24417550 <u>f2</u> 120	1607	6829	63	192	71	0.026	
Protein name				Locu	s Name	Acc#	
				sp:FLI	T_BACSU	P39740	
Description							
FLAGELLAR PROTEIN	FLIT						
			NT	AA			
			1/1 1	~~	Caoro	Drobability	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
ORF Name 24475937_c3_328		<u>AAID</u>			Score	PIODADILITY	
		-		Length	s Name	Acc#	
24475937 <u></u> c3328		-		Length			

	2744474
\$. 31 113	
II.	
113 CE	
pr. 5.00	
T T	
di an	,
羅	
13:11	
A 1141	
200	
ाश क्यंद्र	
51 HS	

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24506692_c1_243	1609	6831		506		
Protein name			<u>  </u>	Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24642837 <u></u> £3 <u></u> 145	. 1610	6832	114	345		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name  24647535f159	-	<u>AAID</u>	Length		Score	Probability
	-		Length	Length	Score Name	Probability  Acc#
24647535_£1_59	-		Length	Length	Ť	* *
24647535_f1_59 Protein name	-		Length	Length	Ť	* *
Protein name  Description	-		Length	Length	Ť	* *
Protein name  Description  NO-HIT	NTID	6833	Length  323  S  NT  Length	Length D72 Locus	Name	Acc#
Protein name  Description  NO-HIT  ORF Name	NTID	AAID	Length  323  NT  Length	Length  Locus  AA  Length	Name	Acc# Probability
24647535_f1_59	NTID	AAID	Length  323  NT  Length	Length  Locus  AA  Length	Name Score	Acc# Probability

			$\underline{\mathbf{NT}}$	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length		
24694187_c1_198	1613 6	835	449 1	350	128	0.00060
Protein name				Locus	s Name	Acc#
lacunin				gp:AF0	78161	AF078161
Description						
Manduca sexta lacu	nin mRNA, co	omplete o	cds.			
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
24864003_f3_131	1614	5836	403 1	212	842	5.2e-84
Protein name				Locus	s Name	Acc#
pantothenate metab			ifp	pir:D6	9878	D69878
homolog yloI:probab		9	1	0		-
1-decarboxylase act Description	:ivase					
Description						
* -			NT	AA	Caoro	Probability
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name 2551.7305_c3_352		<u>AAID</u>	<u>Length</u>		Score	Probability
			Length	Length	Score s Name	Probability  Acc#
25517305_c3_352			Length	Length		
Protein name  Description			Length	Length		
25517305_c3_352			Length	Length		
Protein name  Description			Length	Length	s Name	Acc#
Protein name  Description			Length 455	Length  Locu		Acc# Probability
Protein name  Description  NO-HIT	1615 [	6837	Length  455	Length Locu Locu	s Name	Acc#
Protein name  Description  NO-HIT  ORF Name	1615 [	AAID	Length  455	Length Locu AA Length	s Name	Acc# Probability
25517305_c3_352 Protein name  Description  NO-HIT  ORF Name  25578390_f1_30	NTID	AAID 6838	Length  455  NT Length  418	Length Locu  AA Length Locu  Locu  Locu  Locu	Score  862  S Name	Acc# Probability 4.0e-86
Protein name  Description  NO-HIT  ORF Name  25578390f130	NTID 1616	<u>AAID</u> 6838 se, beta	Length  455  NT Length  418  -subunit	Length  Locu  AA Length  Locu  Locu  gp:PMA	Score  862  S Name  J2015	Acc#  Probability  4.0e-86  Acc#  AJ002015

			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length		
25625438_c1_190	1617	6839	398 1	.197	109	0.0069
Protein name				Locus	3 Name	Acc#
transmembrane pro	tein			gp:YSC	PTM	L11895
Description						
Saccharomyces cer complete cds.	evisiae put	ative tra	nsmembran	e prote	in (PTM1)	gene,
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>		
26210302_f1_10	1618	6840	393	1182	252	3.3e-36
Protein name				Locu	s Name	Acc#
sensory transduct protein slr1983:pr slr1983 Description	ion system cotein slr19	regulator 83:protei	y n	pir:S7	5664	S75664
	NUTT	AAID	<u>NT</u> Length	AA Length	Score	Probability
ORF Name	NTID			480		
26366312 <u></u> c2 <u></u> 28.7	1619	6841	139	±60		» « ·
Protein name						
				Locu	s Name	Acc#
Description				Locu	s Name	Acc#
Description		×	-	Locu	s Name	Acc#
		*	NT		s Name	
	NTID	AAID	NT Length	Locu AA Length	s Name	Acc# Probability
NO-HIT ORF Name		<u>AAID</u>	Length	AA		
NO-HIT			Length	AA Length 405	Score	Probability
NO-HIT  ORF Name  26369000t281	1620	6842	Length	AA Length 405	Score 91 us Name	Probability 0.0067
ORF Name  26369000t281	1620	6842	Length	AA Length 405 Locu	Score 91 us Name	Probability  0.0067  Acc#
ORF Name  26369000t281  Protein name  positive regulate	or for virul	6842 Lence fact	Length  134  cors	AA Length 405 Locu	Score 91 us Name	Probability  0.0067  Acc#  D14877

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
26595260_f2_82	1621 6	843	206 6	21	120	5.9e-06
Protein name				Locus	Name	Acc#
hypothetical protein	n AF0417			pir:A63	9302	A69302
Description						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
26.6.8.7.7.91£3144	1622 6	844	193 5	82		
Protein name				Locus	Name	Acc#
Description						
NO-HIT	<del></del>					
			NT	AA		
ORF Name	NTID	AAID	Length	<u>Length</u>	Score	Probability
26815891 <u>c1</u> 187	1623	845	190 5	73	213	2.4e-17
Protein name				Locus	Name	Acc#
unknown				gp:AF04	8749	AF048749
Description						
Bacteroides fragili	s capsular	polysacc	haride b	iosynthe	sis oper	on,complete
sequence.						
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2767217 <u></u> £2 <u></u> 89	1624 6	846	549 1	650	428	3.3e-39
Protein name				Locus	Name	Acc#
2', 3'-cuclic nucleo	otide 2'-pho	osphodie	sterase	gp:AB02	8630	AB028630
Description						
Clostridium perfring protein, bacterial ho nucleotide 2'-phospho	emoglobin, p	protein-	tyrosine	phosphat	ase, 2',	

100 E	THE THE
2.74	100 100 100 100 100 100 100 100 100 100
H	1212
	N. Tal
	12/22
	727
100	Trans death
115	ner.
Signature Street	i š
	qe,
Trans.	10
ė,	700
#	
ş	11
'n	1112
	222
***	223
Story.	411
	dig
	diff
ij.	##¥
Section 1	172
P.S.	417
	21120

	oility
ORT Name	2-05
2848255_f2_124	
FIOCEIII Hame	Acc#
GIYA gp:AF136495 A	F136495
Description	
Campylobacter lari GlyA (glyA) gene, partial cds.	
NT AA Coore Brobal	oility
ORF Name NTID AAID Length Length	DILLEY
2931557_f1_4	≥-18
Protein name Locus Name	Acc#
probable DNA pol III epsilon chain pir:B71536 B	71536
Description	
ORF Name NTID AAID Length Length	oility
29563916c3345	e-33
Protein name Locus Name	Acc#
galactosyl transferase gp:SPN239004 A	J239004
Description	
Streptococcus pneumoniae type 8 capsular gene cluster.	
ORF Name NTID AAID Length Length	bility
31423265c2276	
J. C. W. C. Nome	Acc#
Protein name Locus Name	
12000211 215	
Description	

ORF Name	NTID A		<u>NT</u> ength <u>L</u>	AA ength	Score	Probability	
31849128_f3_133	1629 685	1 29	89'	7	420	2.7e-39	]
Protein name				Locus	Name	Acc#	
DNA repair protein				pir:A75	391	A75391	
Description							
			NT	AA	Score	Probability	
ORF Name	NTID A	AID Le		ength			_
32040875 <u>f3</u> 132	1630 685	2 26	79	2	394	1.6e-36	_
Protein name					Name	<u>Acc#</u>	
				sp:RECN	_ECOLI		
Description							
DNA REPAIR PROTEIN	RECN (RECOMB)	NATION P	PROTEIN	N)			
ORF Name	NTID A	AAID Le	NT ength	<u>AA</u> Length	Score	Probability	
ORF Name 32681627£13				Length	Score	Probability  2.4e-08	
			ength !	Length  Locus	128 S Name	2.4e-08 <u>Acc#</u>	
32681627 <u></u> £1 <u></u> 3			ength !	Length  Locus	128	2.4e-08	
Protein name  Description	[68]	53 15	ength 1	Length  Locus	128 S Name	2.4e-08 <u>Acc#</u>	
3.26.816.27 <u></u> £1 <u></u> 3 Protein name	[68]	53 15	ength 1	Length  Locus	128 S Name	2.4e-08 <u>Acc#</u>	]
Protein name  Description	, BETA CHAIN,	53 II	ength 1 51 45 TT	Length  Locus	128 S Name	2.4e-08 <u>Acc#</u>	
Protein name  Description  DNA POLYMERASE III	, BETA CHAIN,  NTID	(FRAGMEN	ength 1 51 45 TT	Locus  Sp:DP31  AA  Length	128 s Name B_VIBHA	2.4e-08 <u>Acc#</u> P52620	
Protein name  Description  DNA POLYMERASE III  ORF Name	, BETA CHAIN,  NTID	(FRAGMEN	ength 1	Locus  Sp:DP31  AA  Length	128 s Name B_VIBHA	2.4e-08 <u>Acc#</u> P52620	
Protein name  Description  DNA POLYMERASE III  ORF Name  3322152_c1_199	, BETA CHAIN,  NTID	(FRAGMEN	ength 1	Locus  Sp:DP31  AA  Length	128 S Name S_VIBHA Score	2.4e-08  Acc# P52620  Probability	

Tonoth Tonoth	lity
ORF Name	36
33788882_c1_212 1633 6855 228 687 395 1.2e-	- 30
Protein name Locus Name Ac	<u>cc#</u>
conserved hypothetical protein aq 274 pir:C70325 C70	0325
Description	
NT AA C Duckahi	
ORF Name NTID AAID Length Length Probabi	ility
3.40.70311c1191 1634 6856 347 1044 132 7.0e-	-06
Protein name Locus Name Ac	cc#
transmembrane protein gp:SPAJ6986 AJ	006986
Description	
Streptococcus pneumoniae type 33F DNA, capsular gene cluster.	· .
NT AA Guun Buchahi	
ORF Name NTID AAID Length Length	ility
35784765_t2_61 1635 6857 316 951 593 1.3e-	-57
35.7.84.7.65I.26.1	-57 cc#
Protein name Locus Name A	
Protein name Locus Name A	cc#
Protein name  Locus Name  AU  UDP-N-acetylenolpyruvoylglucosamine reductase gp:BPE238308  AJ	<u>cc#</u> [238308
Protein name  Description  Bordetella pertussis partial gene for putative thioesterase, tRNA-Gly, dapB, omlA genes and partial fur gene.  NT AA Score Probab.	cc# 238308 murB,
Protein name  Description  Bordetella pertussis partial gene for putative thioesterase, tRNA-Gly, dapB, omlA genes and partial fur gene.	cc# 238308 murB,
Protein name  Description  Bordetella pertussis partial gene for putative thioesterase, tRNA-Gly, dapB, omlA genes and partial fur gene.  NT AA Score Probability Score Probab	cc# 238308 murB,
Protein name  Description  Bordetella pertussis partial gene for putative thioesterase, tRNA-Gly, dapB, omlA genes and partial fur gene.  NTID  AAID  NTID  AAID  AAID  Length  Length  Length  Score  Probab: 3914642_f2_80	cc# 238308 murB,
Protein name  Description  Bordetella pertussis partial gene for putative thioesterase, tRNA-Gly, dapB, omlA genes and partial fur gene.  ORF Name  NTID  AAID  NTID  AAID  Length  Length  Length  Score  Probab: 3914642_f2_80	cc# 238308 murB,

•	•		NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	<u>Length</u>	20010	11000011109
3943802_f2_66	1637	6859	133	102	343	4.0e-31
Protein name				Locus	s Name	Acc#
YjgF				gp:AF0	95578	AF095578
Description						
Salmonella typhim	ırium YjgF	(yjgF) gei	ne, compl	ete cds	; and unk	nowngene.
			NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
3944687_£3_143	1638	6860	294	885	221	3.3e-18
Protein name				Locus	s Name	Acc#
hypothetical prote	in AF0417			pir:A6	9302	A69302
Description						0
			NICE	7. 7.		
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4064000 <u>±1</u> _29	1639	6861	88   2	67	83	0.0014
Protein name				Locus	s Name	Acc#
probable integral	membrane pi	rotein	· ×	pir:T3	7050	T37050
Description	×	,				
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
4101561_c2_277	1640	6862	392 1	179		
Protein name				Locus	Name	Acc#
Description					•	
NO-HIT	*		<del></del>	<del></del>		
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
4104636 <u></u> c1192	1641	6863	191 5	76	199	7.2e-16
Protein name			,	Locus	Name	Acc#
serine acetyltrans	ferase			pir:G72	2349	G72349
Description	· · · · · · · · · · · · · · · · · · ·	<del></del>				, <b>Laure</b> ,

The section	1
ş.,	:::7
1.	100
44.	42
Seman.	
\$	1 5
	1890
÷.	11321
ŝ	114:
Share Sa	14
Sir. 18	
382	
22	not,
Strate Street	
Ť,	1
	1
18 AP	100
2	
2	
2	150
報 報報 番	
STATE SHE	
報 報報 番	
報 報報 番	
新 端	PIN N N N N N N N
Trees, & Mark M.	The History and The
清 端	The History and The
新 端	PIN N N N N N N N

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4149067 c3_341	1642	6864	194 5	85	230	3.7e-19	
Protein name			<u> </u>	Locus	Name	Acc#	-
serine acetyltran	isferase			pir:G72	2349	G72349	
Description							
			NT	AA	0	Dwahahilitu	
ORF Name	NTID	AAID	Length	Length	Score	Probability	
4330032_f2_73	1643	6865	536	.611.	135	4.6e-05	]
Protein name				Locus	3 Name	Acc#	
				sp:Y14	3_SYNY3	P74442	
Description							
HYPOTHETICAL WD-1	REPEAT PROTE	IN SLR014	3				
ODE Name			NT	<u>AA</u> Length	Score	Probability	
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
4690675 <u></u> £3 <u></u> 152	NTID		<u>NT</u> Length	Length	Score s Name	Probability  Acc#	
46.906.75f3152 Protein name	NTID	AAID	<u>NT</u> Length	Length		-	
Protein name  Description	NTID	AAID	<u>NT</u> Length	Length		-	
46.906.75f3152 Protein name	NTID	AAID	NT Length 225	Length 578 Locu		-	
Protein name  Description  NO-HIT	NTID	AAID	<u>NT</u> Length	Length		-	
Protein name  Description	NTID 1644 NTID	<u>AAID</u> 6866	NT Length 225	Length  Locu	s Name	Acc#	
Protein name Description NO-HIT ORF Name	NTID 1644 NTID	AAID 6866 AAID	NT Length 225 (	Length Locus  AA Length	s Name	Acc# Probability	
### 4690675f3152  Protein name  Description  NO-HIT  ORF Name  ###################################	NTID 1644 NTID	AAID 6866 AAID	NT Length 225 (	Length Locus  AA Length Locus Locus Locus	Score	Acc# Probability  5.2e-74	
### 4690675f3152  Protein name  Description  NO-HIT  ORF Name  ###################################	NTID 1644 NTID	AAID 6866 AAID	NT Length 225 (	Length Locus  AA Length Locus Locus Locus	Score 756 S Name	Acc#  Probability  5.2e-74  Acc#	

ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
4773261_c1_196  Protein name	1646	6868	237 7	14 Locus	103 Name	0.017 Acc#	
				sp:YJBH	ECOLI	P32689	
Description						-	
PRECURSOR							
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
4802168_c3_406	1647	6869	206	21			
Protein name				Locus	: Name	Acc#	
Description						• •	
NO-HIT							
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability	
ORF Name 4882192_c1_194	NTID 1648	<u>AAID</u> 6870	Length		Score	Probability	
			Length	Length		2.184	
4882192 <u></u> c1 <u></u> 194	1648	6870	Length	Length	162 Name	1.4e-09	
Protein name  probable lipopolysa	1648	6870	Length	Length  Locus	162 Name	1.4e-09 <u>Acc#</u>	
Protein name  probable lipopolysa N-acetylglucosaminyl	1648	6870	Length	Length  Locus	162 Name	1.4e-09 <u>Acc#</u>	]
Protein name  probable lipopolysa N-acetylglucosaminyl  Description	1648 ccharide transfera	se, rfbU	Length  409 [1]	Length  Locus  pir:F64	162 S Name	1.4e-09 Acc# F64500	
Protein name  Probable lipopolysa N-acetylglucosaminyl  Description  ORF Name	ccharide transfera	se, rfbU	Length  409 [1]	Length  Locus  pir:F64  AA  Length	Name 1500 Score	Acc# F64500  Probability	
Protein name  Probable lipopolysa N-acetylglucosaminyl  Description  ORF Name  4898450_t2_62	ntide	se, rfbU	Length  409 [1]	Length  Locus  pir:F64  AA  Length	Score 458 Name	1.4e-09 Acc# F64500  Probability 2.6e-43	

$\frac{\mathrm{NT}}{\mathrm{NT}}$ Score Probability
ORF Name NTID AAID Length Length
4954380_t2_90 [6872] [620] [1863] [357] [4.4e-47]
Protein name Locus Name Acc#
oxaloacetate decarboxylase, subunit alpha pir:C69406 C69406 (oadA) homolog
Description
NT AA Score Probability ORF Name NTID AAID Length Length
4957965f31301651
Protein name Locus Name Acc#
sp:DP3B_PSEPU P13455
Description
DNA POLYMERASE III, BETA CHAIN,
NT AA Dackehilitus
Score Probability
ORF Name NTID AAID Length Length
ORF Name NTID AAID Length Length
ORF Name         NTID         AAID         Length         Length           50.86.53.7_f2_76
ORF Name         NTID         AAID         Length         Length           5.0.8.6.5.3.7t27.6
ORF Name         NTID         AAID         Length         Length           5.0.8.6.5.3.7t27.6
ORF Name  NTID AAID Length Length  5.0.86.53.7t27.6
ORF Name  NTID  AAID  Length  Length  SOR6.53.7_f.276
ORF Name  NTID  AAID  Length  Length  Length  S.086.53.7_t276
ORF Name  NTID AAID Length Length    Source   So
ORF Name  NTID AAID Length Length    Length   Length
ORF Name  NTID AAID Length Length    Source   So

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
	1654	6876		999	443	1.0e-41	٦
	1031		L L		s Name	Acc#	_
Protein name							
ss-1,4-galactosyltra	ansferase			gp:SPC	PS14E	X85787	
Description							
S.pneumoniae cps14	locus.						
			NT	AA			
ORF Name	NTID	AAID	Length	<u> AA</u> Length	Score	Probability	
6048452_£1_21	1655	6877	78	237			
Protein name				Locu	s Name	Acc#	
Description							
NO-HIT							
			NT	AA	Caoro	Drobability	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
ORF Name  6.4146.77_c1_188	NTID	<u>AAID</u> 6878	Length		Score	Probability	
			Length	Length			
Protein name  NADH dehydrogenase	1656	6878	Length	Length	149	1.5e-07	
6.4146.77_cl_188 Protein name	1656	6878	Length	Length 1218 Locu	149	1.5e-07 <u>Acc#</u>	
Protein name  NADH dehydrogenase	1656	6878	Length	Length 1218 Locu	149	1.5e-07 <u>Acc#</u>	
Protein name  NADH dehydrogenase subunit homolog	1656	6878	Length 405	Length  1218  Locu	149	1.5e-07 Acc# H69478	
Protein name  NADH dehydrogenase subunit homolog	1656	6878	Length	Length 1218 Locu	149	1.5e-07 <u>Acc#</u>	
Protein name  NADH dehydrogenase subunit homolog  Description  ORF Name	(ubiquino	ne),, 39	Length 405	Length  Locu  pir:H6	149 s Name	1.5e-07 Acc# H69478	
Protein name  NADH dehydrogenase subunit homolog  Description	(ubiquino	6878 ne),, 39	Length  405  kDa  NT  Length	Length  Locu  pir:H6  AA  Length	149 IS Name 9478 Score	1.5e-07 Acc# H69478	
Protein name  NADH dehydrogenase subunit homolog  Description  ORF Name  6694425_63_365	(ubiquino	6878 ne),, 39 <u>AAID</u> 6879	Length  405  kDa  NT  Length	Length  Locu  pir:H6  AA  Length	149 IS Name 9478  Score 420 IS Name	1.5e-07 Acc# H69478  Probability 2.7e-39	

ORF Name	NTID	AAID	<u>NT</u> <u>Length</u>	AA Length	Score	Probability
6742762_f1_28	1658	6880	61	186		
Protein name				Locus	Name	Acc#
Description						
NO-HIT						
			NT	7.7		
ORF Name	NTID	AAID	Length	<u>AA</u> Length	Score	Probability
6.83.4387±119	1659	6881	120	363		
Protein name				Locus	Name	Acc#
Description				*		
NO-HIT						-
			NТ	ΔΔ		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name 6845277c2288		<u>AAID</u>	Length		Score	Probability
		-	Length	Length	Score  S Name	Probability  Acc#
6845277 <u></u> c2 <u></u> 288		-	Length	Length		
6.845277_c2_288 Protein name		-	Length	Length		
Protein name  Description		-	Length	Length  195  Locus		Acc#
Protein name  Description		-	Length	Length		
Protein name  Description  NO-HIT	MTID	6882	Length  64  NT  Length	Length  Locus	: Name	Acc#
Protein name  Description  NO-HIT  ORF Name	MTID	6882 AAID	Length  64  NT  Length	Length  Locus  AA  Length	Score	Acc# Probability
Protein name  Description  NO-HIT  ORF Name  819433_c3_340	NTID	AAID	NT Length 371	Length  Locus  AA  Length	Score 339 Name	Acc# Probability 1.0e-30

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Lenqth	Score	Probability
960825 ±2 69	1662	6884		398	771	1.7e-76
		0001	100			Acc#
Protein name					Name	
phosphate starvat	cion inducib.	Le protei	1	pir:A69	9873	A69873
Description				1		
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
9773281f2125	<del></del>	6885		210		
					s Name	Acc#
Protein name				<u>1000</u>	3 Name	210011
Description						
NO-HIT						× ×
			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	NT Length	<u>AA</u> Length	Score	Probability
ORF Name  10573830c1296	*	AAID 6886	Length		Score	Probability
	*		Length	Length	Score S Name	Probability  Acc#
105.73830 <u></u> c1296	*		Length	Length		-
1.05.73.830c1296 Protein name	*		Length	Length		-
Protein name  Description	*		Length	Length  Locus	s Name	Acc#
Protein name  Description	*		Length	Length		-
Protein name  Description  NO-HIT		6886	Length  168  NT Length	Length Locus	s Name	Acc#
Protein name  Description  NO-HIT  ORF Name		AAID	Length  168  NT Length	Length  Locus  AA  Length	Score	Acc# Probability
Protein name  Description  NO-HIT  ORF Name  10664128c3491	NTID	AAID	Length  168  NT Length	Length  Locus  AA  Length	Score 125 S Name	Acc# Probability  1.8e-11
Protein name  Description  NO-HIT  ORF Name  10664128_c3_491  Protein name	NTID	AAID	Length  168  NT Length	Length  Locus  AA  Length  645  Locu	Score 125 S Name	Acc#  Probability  1.8e-11  Acc#

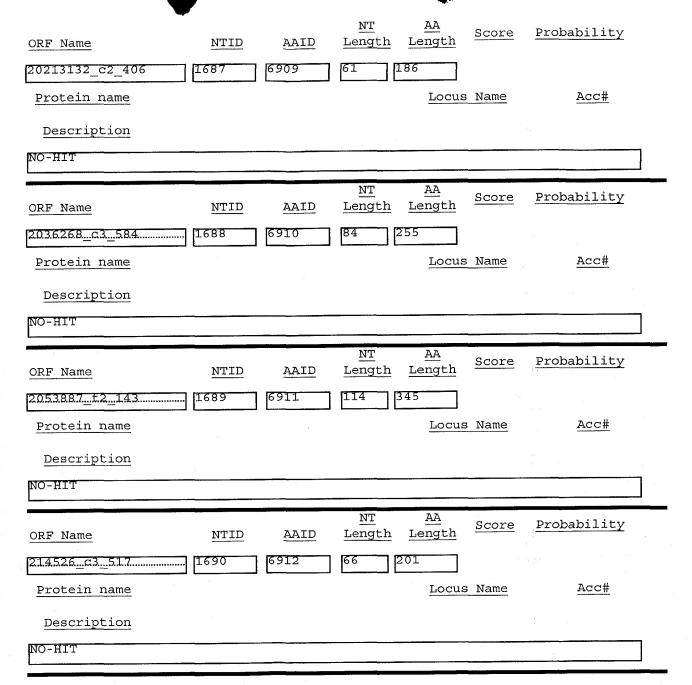
·			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
10720337_f1_75	1666	6888	211	36	258	4.0e-22
Protein name				Locus	Name	Acc#
				sp:YC08	YEAST	P37261
Description						
HYPOTHETICAL 21.1	KD PROTEIN	IN FUS1-A	AGP1 INTI	ERGENIC R	EGION	
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	<u>30016</u>	FIODADITICY
10986288_f3_169	1667	6889	406	1221	149	2.2e-07
Protein name				Locus	Name	Acc#
hypothetical prot	ein BBI16			pir:G70	0241	G70241
Description						
			NT	AA	~	D1-1-1-1-1-1-
ORF Name	NTID	AAID	Length	Length	Score	Probability
10988261 <u></u> f1 <u></u> 7	1668	6890	714	2145	1016	1.9e-102
Protein name				Locus	<u>Name</u>	Acc#
DNA topoisomerase	III topB			pir:H69	724	H69724
Description						
	· · · · · · · · · · · · · · · · · · ·		NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability
11910250 <u></u> £2 <u></u> 102	1669	6891	159	480		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT						
			NT	AA		
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability
11924205 <u></u> c3539	1670	6892	110	333		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT						

ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
1212762_c2_390	1671	6893	300	03	354	2.7e-32
Protein name				Locus	s Name	Acc#
				gp:AB0	L2957	AB012957
Description						
Vibrio cholerae g	enes for o-a	antigen s	ynthesis,	strain	022, com	mpletecds.
			NT	AA	-	Desch alad I desc
ORF Name	NTID	AAID	Length	Length	Score	Probability
12298425_c2_391	1672	6894	300	03	219	5.5e-18
Protein name				Locus	s Name	Acc#
putative glycosyl	transferase	3		gp:AF0	18749	AF048749
Description						
Bacteroides fragi	lis capsula:	r polysac	charide h	piosynthe	esis oper	con, complete
sequence.						
ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
12501087±2116	1673	6895	168	507		
Protein name				Locus	s Name	Acc#
Description						
						-
NO-HIT					*	
			NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	-	
12691280 <u></u> c3510	1674	6896	1022	3069	726	1.3e-81
Protein name				Locu	s Name	Acc#
probable swf/snf	helicase			pir:E7	1481	E71481

		71 77 77 75	NT	AA Longth	Score	Probability
ORF Name	NTID	AAID	<u>Length</u>	Length		
13885212_f3_220	1675	6897	388	167		
Protein name				Locus	8 Name	Acc#
Description						
NO-HIT						
			NT	AA	-	D
ORF Name	NTID	AAID	Length	Length	Score	Probability
14116635 <u></u> £2 <u></u> 158	1676	6898	416	1251		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT						
		<del>-</del>	NT	AA		
ORF Name	NTID	AAID	Length	Length	Score	Probability
14736262 <u></u> c3 <u></u> 545	1677	6899	427	L284	241	7.6e-20
Protein name				Locus	s Name	Acc#
MocB (Tn4399)				pir:B4	8487	B48487
Description	-					
	-		NT	AA	Score	Probability
ORF Name	NTID	AAID	Length	Length	•	-
15017287_f3_252	1678	6900	837	2514	391	2.2e-32
Protein name				Locu	s Name	Acc#
enhanced entry prot	ein EnhC		-	gp:AF0	57704	AF057704
Description			*	1	*	
Legionella pneumoph EnhC (enhC) genes, o			nhB (enhl	B) , and $G$	enhanced	entryprotein

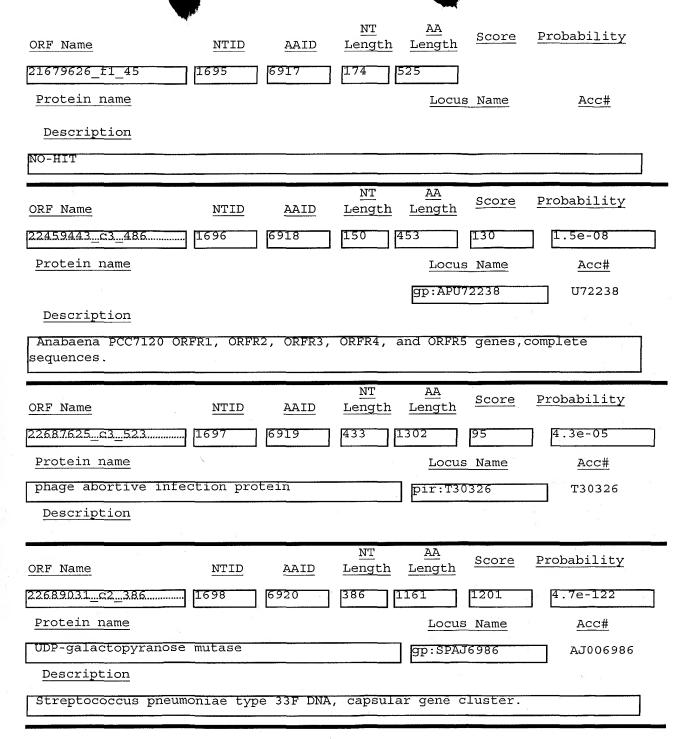
			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length		
16135886_f3_193	1679	6901	103	312		
Protein name				Locus	Name	Acc#
m						
Description						
NO-HIT						
	··		NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length		
16229025_f2_124	1680	6902	280	343	105	0.00081
Protein name				Locus	Name	Acc#
				sp:YS21	BORBU	. *
Description						•
	יאי החחמו					
HYPOTHETICAL PROTEI	TZGGG N.				1	
				· ·		
			NT	<u>AA</u>	Score	Probability
ORF Name	NTID	AAID	Length	Length		
ORF Name	NTID	<u>AAID</u>	Length		<u>Score</u>	Probability
			Length	Length		
16829717 <u></u> £3 <u></u> 188			Length	Length	1011 S Name	1.5e-118
16829717t3188 Protein name			Length	Length 5856 Locus	1011 S Name	1.5e-118
Protein name  Description	1681	6903	Length	Length  5856  Locus  gp:AB0	1011 S Name	1.5e-118 Acc#
16829717t3188 Protein name	1681	6903	Length	Length  5856  Locus  gp:AB0	1011 S Name	1.5e-118 Acc#
Protein name  Description  Agrobacterium tumes	1681	6903	Length  [1951]  -SAKURA,	Length  5856  Locus  gp:AB03  complete	1011 S Name	1.5e-118 Acc#
Protein name  Description  Agrobacterium tumes  ORF Name	1681 Eaciens pl	asmid pTi	Length  1951  -SAKURA,  NT Length	Length  Locus  gp:AB0  complete  AA  Length	1011 S Name 16260 Score	1.5e-118  Acc#  Ce.  Probability
Protein name  Description  Agrobacterium tumes	1681	6903	Length  1951  -SAKURA,  NT Length	Length  5856  Locus  gp:AB03  complete	1011 S Name 16260 S sequence	1.5e-118  Acc#  Dece.  Probability  2.6e-176
Protein name  Description  Agrobacterium tumes  ORF Name	1681 Eaciens pl	asmid pTi	Length  1951  -SAKURA,  NT Length	Length  Locus  gp:AB0  complete  AA  Length	1011 S Name 16260 Score	1.5e-118  Acc#  Ce.  Probability
Protein name  Description  Agrobacterium tumes  ORF Name  16832885_f2_168	1681  Eaciens pl  NTID  1682	asmid pTi	Length  1951  -SAKURA,  NT Length	Length  Locus  gp:AB0  complete  AA  Length	1011 S Name 16260 Score 1713 S Name	1.5e-118  Acc#  Dece.  Probability  2.6e-176

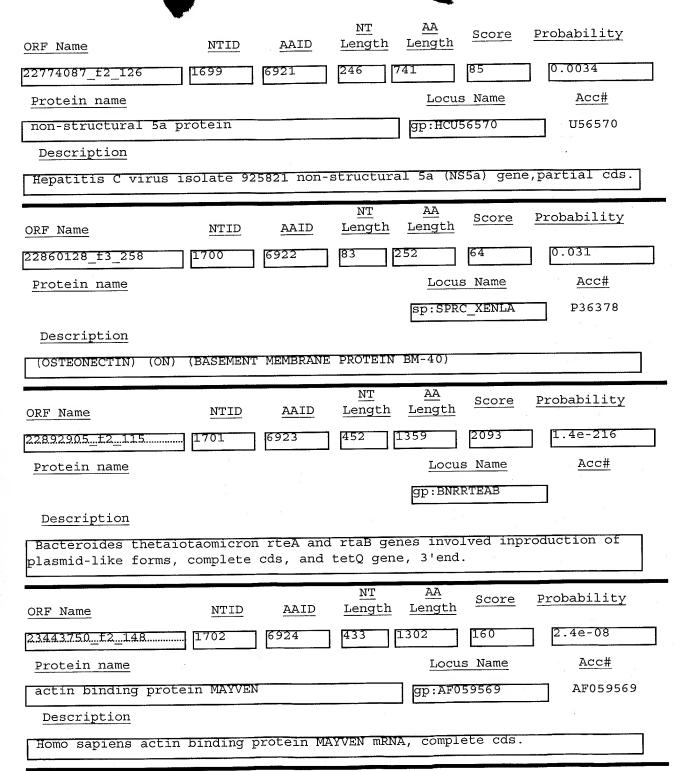
ORF Name	NTID	AAID	NT AA Score Probability Length Length
19562660_£2_134	1683	6905	478 1437 113 0.0074
Protein name			Locus Name Acc#
ES/130			gp:AF006751 AF006751
Description			
Homo sapiens ES/130	mRNA, co	mplete cd	S.
ORF Name	NTID	AAID	NT AA Score Probability
197131_f1_35	1684	6906	335 1008
Protein name			Locus Name Acc#
Description			
NO-HIT			
			NT AA Guun Buchahilitus
ORF Name	NTID	AAID	Length Length Score Probability
19945402 <u>f1</u> 57	1685	6907	375 1128
Protein name			Locus Name Acc#
Description			
NO-HIT			
ORF Name	NTID	AAID	NT AA Score Probability Length Length
1995941£289	1686	6908	427   1284   1049   6.1e-106
Protein name			Locus Name Acc#
transposase			gp:AF038866 AF038866
Description			Y Y
Bacteroides fragili protein BmpH (bmpH)			0 transposase (bipH) andmobilization s.





ORF Name	NTID	AAID	<u>NT</u> Length	AA Length	Score	Probability
21484662_f1_6	1691	6913	532	.599	144	9.4e-07
Protein name				Locus	s Name	Acc#
				sp:M49_	STRPY	P16947
Description						
M PROTEIN, SEROTY	PE 49 PRECU	RSOR				
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
21515632_f1_17	1692	6914	785	2358	3809	0.0
Protein name				Locus	s Name	Acc#
tetracycline resi RteA	stance eleme	ent regula	ator	pir:A4	1860	A41860
Description					-	
ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
2152643.7 <u></u> £2 <u></u> 13.0	1693	6915	138	117		
Protein name				Locus	s Name	Acc#
Description						
NO-HIT				-	*	*
ODE Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability
ORF Name		6916		102		
216.016.25c1295 Protein name	1094	0310			s Name	Acc#
Description						
NO-HIT				0		





ORF Name	NTID	AAID	<u>NT</u> Length	<u>AA</u> Length	Score	Probability	
23492786_f2_162	1703	6925	85	258			
Protein name			L	Locus	Name	Acc#	
Description							
NO-HIT							
			NT	AA			_
ORF Name	NTID	AAID	<u>Length</u>	<u>Length</u>	Score	Probability	
23596911 <u></u> t2 <u></u> 139	1704	6926	77	234			
Protein name				Locus	Name	Acc#	
Description							
NO-HIT			Ø				
			NT	AA			_
ORF Name	NTID	AAID	<u>Length</u>	Length	Score	Probability	
23644552 <u></u> £1 <u></u> 31	1705	6927	941	2826	537	8.3e-85	]
23644552f131 Protein name	1705	6927	941		537 Name	8.3e-85 Acc#	]
	1705	6927	941 [2		Name		
	1705	6927	941	Locus	Name	Acc#	
Protein name			) a	Locus gp:BFU6	Name	Acc#	
Protein name  Description			plete cds	Locus gp:BFU6	3096	Acc# U63096	
Protein name  Description			) a	Locus gp:BFU6	Name	Acc#	
Protein name  Description  Bacteroides fragili	s (bctA)	gene, com	plete cds  NT  Length	Locus gp:BFU6	3096	Acc# U63096	]
Protein name  Description  Bacteroides fragili  ORF Name	s (bctA)	gene, com	plete cds  NT  Length	Locus gp:BFU6  AA Length	Name 3096 Score	Acc# U63096  Probability	
Protein name  Description  Bacteroides fragili  ORF Name  23.710.777cl327	.s (bctA)  NTID  1706	gene, com	plete cds  NT  Length	Locus gp:BFU6  AA Length	Score  81 Name	Acc# U63096  Probability  0.029	]
Protein name  Description  Bacteroides fragili  ORF Name  23710777_cl_327	.s (bctA)  NTID  1706	gene, com AAID	plete cds  NT  Length	Locus gp:BFU6  AA Length Locus	Score  81 Name	Acc# U63096  Probability  0.029	]





ORF Name	NTID	AAID	Length	<u>AA</u> Length	Score	Probability	
24251937_t1_1	1707	6929	94	285			
Protein name				Locus	Name	Acc#	
Description						•	
NO-HIT							]
			NT	AA			
ORF Name	NTID	AAID	Length	Length	Score	Probability	
24259442 <u></u> £3 <u></u> 246	1708	6930	313	942	370	5.4e-34	
Protein name				Locus	Name	Acc#	
				sp:GSP/	_BACSU	P25148	
Description							
GENERAL STRESS PROT	EIN A						
							l
			NT	ΔΔ			
ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability	
	NTID 1709	<u>AAID</u>	Length		Score	Probability	
ORF Name			Length	Length	Score S Name	Probability  Acc#	
ORF Name  243470.90E134			Length	Length			
ORF Name  24347090f134  Protein name			Length	Length			
ORF Name  24347090f134  Protein name  Description			Length	Length 597 Locus			
ORF Name  24347090f134  Protein name  Description			Length	Length			
ORF Name  24347090f134  Protein name  Description  NO-HIT	1709	6931	Length  198  NT  Length	Length  597  Locus	s Name	Acc#	
ORF Name  243470.90f134  Protein name  Description  NO-HIT  ORF Name	1709 NTID	AAID	Length  198  NT  Length	Length  Locus  AA  Length	Score	Acc#	* * *
ORF Name  243470.90f134  Protein name  Description  NO-HIT  ORF Name  244410.812f158	1709 NTID	AAID	Length  198  NT  Length	Length  Locus  AA  Length	Score	Acc# Probability	



